Key Elements for Educational Accountability Models

Prepared by
Marianne Perie, Center for Assessment
with
Judy Park, Utah
Kenneth Klau, Massachusetts

A paper commissioned by the
Council of Chief State School Officers
Accountability Systems and Reporting State Collaborative

December 2007

Council of Chief State School Officers
Washington, DC

COUNCIL OF CHIEF STATE SCHOOL OFFICERS

The Council of Chief State School Officers (CCSSO) is a nonpartisan, nationwide, nonprofit organization of public officials who head departments of elementary and secondary education in the states, the District of Columbia, the Department of Defense Education Activity, and five U.S. extra-state jurisdictions. CCSSO provides leadership, advocacy, and technical assistance on major educational issues. The Council seeks member consensus on major educational issues and expresses their views to civic and professional organizations, federal agencies, Congress, and the public.

STATE COLLABORATIVE ON ASSESSMENT AND STUDENT STANDARDS

The State Collaborative on Assessment and Student Standards (SCASS) Project was initiated in 1991 to encourage and assist States in working collaboratively on assessment design and development of high quality assessments in relation to standards for student learning. State education agencies voluntarily support collaborative work across a variety of topics and subject areas. The Accountability Systems and Reporting (ASR) state collaborative project has been assisting States to develop and improve statewide systems since 2000. Currently 28 state education agencies are participating members, together with researchers, consultants, and measurement specialists. Priority activities of the ASR Collaborative are: state planning and development to meet requirements for state systems under NCLB, sharing strategies and research-based evidence to improve the validity of accountability systems, developing reporting formats that improve information access for constituents, and establishing standards for the essential components of accountability systems. Funding support for this paper was entirely from member States of the Accountability Systems and Reporting State Collaborative (ASR-SCASS). For further information about ASR SCASS and resources on accountability across the 50 States, see the CCSSO web site: http://www.ccsso.org/projects/Accountabilty_Systems/.

2007

Council of Chief State School Officers
Elizabeth Burmaster (Wisconsin), President
Rick Melmer (South Dakota), President-Elect
Valerie Woodruff (Delaware), Past President
Gene Wilhoit, Executive Director

Rolf K. Blank, Director of Education Indicators

Council of Chief State School Officers
Attn: Publications
One Massachusetts Ave., NW, Suite 700
Washington, DC 20001
202-336-7016
Fax: 202-408-8072
www.ccsso.org

ISBN 1-884037-24-0

Key Elements for Educational Accountability Models

Prepared by
Marianne Perie, Center for Assessment

with
Judy Park, Utah
Kenneth Klau, Massachusetts

A paper commissioned by the
Council of Chief State School Officers
Accountability Systems and Reporting State Collaborative

December 2007

The authors wish to thank all members of the Accountability Systems and Reporting State Collaborative for their assistance in conceiving and executing this paper, especially the following:

Jacqueline Soychak, Maine
Rachelle Tome, Maine
Pat McCabe, California
Pat Roschewski, Nebraska
Kenna Seal, West Virginia
Mitchell Chester, Ohio
Robin Taylor, Delaware
Margie Vandeven, Missouri
Jennifer Stegman, Oklahoma
Kevin Hill, Kentucky
Cathy Wagner, Minnesota
Bob Bernstein, California
Rolf Blank, CCSSO
Andra Williams, CCSSO

We also wish to thank our expert reviewers for their advice and feedback, Ellen Forte, Bill Erpenbach, Arie van der Ploeg, Bob Linn, and Sue Rigney.

Table of Contents

Appendices

List of Figures

Executive Summary

The decision to work toward this report was stimulated by leaders of state departments of education asking the question "where can I go to find key information about developing state accountability systems?" Currently, there is no one document that provides a set of recommendations based on evidence or experience that covers all of the decisions involved with developing state accountability systems, and at the same time there is an extensive list of papers, articles, and other resources that should be read and considered. The educational measurement and evaluation field has spent years developing its latest version of *Standards for Educational and Psychological Testing* (APA, AERA & NCME, 1999) and plans to spend several years developing the next version. This resource has become a common, much-used resource to many in the testing industry to support inferences made from tests or to critique testing practices. However, there is no parallel document for educational accountability systems. Although several organizations and individual researchers have recommended practices for accountability systems, we have not yet collaborated as a field to produce a universally accepted set of standards.

The purpose of this report is to summarize the work that has been done to date on developing a set of standards for accountability and inform those not familiar or well experienced in accountability about essential elements of a good/valid accountability system. In addition, we wanted to create a tool that states could use in developing a new accountability system or in evaluating a current one. The audience is intended to be state or district policymakers who are designing, redesigning, or reviewing their accountability systems. This report expands accountability beyond the federal definition under NCLB. Thus the overarching goal of this report is to answer the question "As states consider developing accountability plans separate from NCLB or as an enhancement to NCLB, what are the elements, standards, etc. of a good accountability system?"

To accomplish this goal, the Accountability Systems and Reporting State Collaborative (ASR SCASS), managed by the Council of Chief State School Officers, developed a process for working with state leaders on identifying the key elements for an accountability system, and ASR selected the National Center for the Improvement of Educational Assessment to serve as consultants and writers for the project, including work on three tasks:

1. Synthesize the literature to date on critical elements in an accountability system
2. Develop a framework incorporating core concepts, guiding questions, key elements, and various options listed in previous literature
3. Provide descriptive examples of more fine-grained elements or options that increase effectiveness and success of state-level accountability systems

As the report was developed, three complementary documents were created. After reviewing the immense amount of literature already published on accountability systems, both from education and business contexts, an annotated bibliography was developed summarizing over 70 reports, articles, and chapters. A second product is a matrix developed to synthesize all of the reviewed literature. It was the first step to developing our framework of core concepts, guiding questions, key components, and options discussed in this paper. It is available in full color with each color referencing a specific article, chapter, or report. The guiding questions

or key components are written in the author's own words to allow the reader to trace any feature back to its original source.

Finally, the third product is a checklist for designing or evaluating an accountability system. The checklist is organized by core concept, and each relevant section is presented as in introduction to a chapter, also organized by core concept. It was created to help state- and district-level policymakers evaluate the design of their accountability system. A checklist approach seemed most user-friendly as it highlights the considerations states should undertaking during either the development of a new accountability system or an evaluation of a current one.

The checklist is organized under seven core concepts based on the various organizational strategies of leading researchers:

1. Goals
2. Performance indicators
3. Design decisions
4. Consequences
5. Communication
6. Support
7. System evaluation, monitoring, and improvement

Within each of these core concepts are a list of guiding questions and a set of key elements that provide guidance on developing a strong educational accountability system. This report is organized around those seven questions, with the key elements described and explained in each chapter. Examples from state accountability systems also serve to provide rich detail regarding the various components.

Chapter 1: Introduction and Background

Educational accountability has been a much-used phrase since the 1970s. The 1980s saw a movement to standards-based accountability. The Improving America's Schools Act (IASA) of 1994 moved the discussion to state-level educational accountability systems. The onus was initially put on state governments to define their accountability systems. The *No Child Left Behind Act of 2001* (NCLB) laid out a much more prescriptive accountability system, providing a specific framework within which states must develop their accountability system. Because of this legislation, the field as a whole has learned a great deal about education accountability. Much has been written about what works and what does not. Yet there still does not exist one comprehensive set of standards or key elements for developing effective educational accountability systems.

The educational measurement and evaluation field has spent years developing its *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 1999) and plans to spend several years developing the next version. The Testing Standards document has become invaluable as the common reference source for many decisions and practices in the testing industry, and it is typically cited to support inferences made from tests or to critique testing practices. Likewise, resources exist that provide standards for student and program evaluation (Joint Committee on Standards for Education Evaluation, 2002 & 2004). In addition, the American Evaluation Association has produced documents guiding evaluators and those using high-stakes tests for evaluative purposes (AEA, 2002). However, there is no parallel document for educational accountability systems. Although several organizations and individual researchers have created "standards" for accountability system, we have not yet collaborated as a field to produce a universally accepted set of standards.

This report was prepared by the ASR State Collaborative in response to requests from state department of education leaders focused on the question "where can I go to find key information about developing state accountability systems?" The answer, it turned out, was a list of resources that were all "must reads." The present report is intended to synthesize and organize the available information on accountability research and practices and assemble it all into one document. The "key elements" identified in the report are intended to clarify and document what must be present in an effective accountability system.

Defining Accountability

The first step for any book or article on education accountability is to define accountability. However, defining accountability has become more complex as our understanding of accountability has grown. In the past, definitions focused primarily on the interaction of goals, indicators, decision rules, and consequences. Although those components are still central to any accountability system, the latest movement has been to focus more on capacity building and providing appropriate supports. That is, the purpose of accountability is not simply to identify and punish ineffective schools, but to provide appropriate supports to ensure that all schools are effective.

To illustrate, consider the following two definitions proposed in the 1970s and 1990s:

> An accountability system is a closed loop reflecting a chain of responses to perceived needs or demands; an activity or set of activities that emerges to fill those demands; outcomes that result from those activities; and feedback on outcomes to the source of the demands. The feedback may generate new demands or a regeneration of the old ones; in either case, the previous set of activities may be modified or remain intact; a new or altered set of activities may be modified or remain intact; a new or altered set of outcomes may be produced; and the loop is completed again with feedback to the source of the demands (Levin, 1974, p. 375).

> Accountability describes a relationship between two parties in which four conditions apply: first, one party expects the other to perform a service or accomplish a goal; second, the party performing the activity accepts the legitimacy of the other's expectation; third, the party performing the activity derives some benefits from the relationship; and fourth, the party for whom the activity is performed has some capacity to affect the other's benefits (Stecher and Hanser, 1992).

Both definitions imply a reciprocal relationship between results and positive or negative consequences. Accountability systems start with a set of goals and a theory of action that states that a specific act will produce a desired outcome. Those actions are rewarded when successful; other actions that do not produce the desired outcome are sanctioned. The system must undergo constant monitoring to ensure that the action will produce the desired outcome, that the rewards and sanctions are effective, and that the feedback given to the various parties provides useful information on how to adapt their actions to produce the desired outcomes.

However, neither definition focuses on the responsibility of the party who sets the goals to provide the supports for meeting those goals. Accountability provides a mechanism for thinking globally about schooling, how we measure the effectiveness of our schools in promoting learning, and how we provide the supports necessary to ensure our schools become even more effective. To that end, there has also been a greater recognition of the reciprocity of the relationship between state governance, district governance, and school governance in that each has both expectations of and responsibilities to the others. There has been more work exploring the essential conditions (not just structures) that must be in place to effect change in low-performing schools, such as time for teachers to collaborate, aligned curriculum, data systems that provides useful data, and coaches for reading and math. Ultimately, accountability has become more than a system to evaluate schools and allocate sanctions and rewards: it is a system that allows the public to understand how well their schools are working and to provide information to policymakers on the changes that are needed to make the schools more effective and to continually improve all students' educational opportunities.

Brief History of Accountability

Looking back over the history of accountability, the approach to designing accountability systems has changed over time. The field of education has moved from financial accountability and accountability based on inputs to standards-based accountability systems based on outputs. Fuhrman (1999) focuses on how the goals of accountability models have moved

from compliance with regulations to improved student performance with a resulting focus on outcome data such as test scores and graduation results. Goertz (2001) likewise discusses some of the changes in educational accountability over the past 30 years, including whether the federal education requirements supplement or supplant state and district requirements. She also discusses changes from procedural accountability to educational accountability.

In addition to changes in the type of accountability measures, other changes are occurring in terms of who is accountable. Accountability systems are now focusing more on schools as the unit of change. As a result, continuous improvement strategies are being adopted at the school level. Another area of change is how results are reported. For example, more categories of performance are being developed to report student scores. Students are no longer tested as passing or failing but often receive information that places them in one of four to five performance levels (e.g., Below Basic, Basic, Proficient, and Advanced). Schools are often labeled as either effective or needing improvement. Finally, more consequences are attached to student performance, including public reporting and tangible forms of recognition. Poor performance is now likely to result in sanctions such as state intervention, visits by technical advisory teams, and reconstitution of schools.

Student assessment for educational accountability has moved from matrix sampling to whole system. That is, in the past, schools, teachers, or students were matrix sampled so that only a sample was given tests or other instruments to measure the desired outcomes. Then these results were generalized to the full population. Now, however, the movement has been towards assessing every student and evaluating every teacher and school.

The recent changes in educational accountability have resulted in some state systems that have become overloaded in trying to serve too many purposes simultaneously, failing to serve any of them well. In addition, expectations have increased. Stakeholders expect to see immediate improvements and evidence that student learning has increased, leading to quick fixes or short-term strategies. The changes have discouraged a thoughtful approach to the entire educational accountability system focused on the goals.

Overview of the Report

The report focuses on a whole-system standards-based approach to educational accountability and the report is designed to provide information about essential elements of a good/valid accountability system. A central feature of this report is a tool that states can use in developing a new accountability system or in evaluating a current one. It is intended for use by state or district policymakers who are designing, redesigning, or evaluating their accountability systems. The audience may include legislators, state and district superintendents, state or local boards of education, directors of accountability and assessment, those providing supports and training to schools, school principals and education leaders, as well as researchers. This report is not intended to discuss accountability systems only under the NCLB system, but beyond the federal definition. Thus the overarching goal of this report is to answer the question "As states consider developing accountability plans separate from NCLB or as an enhancement to NCLB, what are the elements, standards, etc. of a good accountability system?"

To accomplish this goal, the Accountability Systems and Reporting SCASS contracted with the National Center for the Improvement of Educational Assessment to undertake three tasks toward completing the paper:

1. Synthesize the literature to date on critical elements in an effective accountability system
2. Craft a framework incorporating core concepts, guiding questions, key elements, and various options listed in previous literature
3. Provide descriptive examples of more fine-grained elements or options that increase effectiveness and success of state-level accountability systems

The results of the first two steps are summarized in the next section. We developed a framework to use as a supporting organizational structure, a matrix and checklist created to examine accountability systems using a series of questions and key elements within several major topics. The framework was developed through a synthesis of the current literature and resulted in seven core concepts fundamental to any educational accountability system. Each of these core concepts is discussed in its own chapter. Each chapter provides additional detail from the literature review, focuses on the guiding questions, key elements, and options within that core concept, and provides descriptive examples from various state departments of education. These examples, as described in step 3 above, are illustrations of how different components have been implemented in one or more states.

There are also three appendices to this report. The first appendix includes a full bibliography, annotated with 2-3 sentence summaries each article, report, chapter, or book. This bibliography matches the list of recommended reading at the end of each chapter. The second appendix includes the full matrix developed to synthesize all of the literature. It was the first step to developing our framework of core concepts, guiding questions, key elements, and options. It is presented in color, with each color referencing a specific article, chapter, or report. The guiding questions or key components are written in the author's own words to allow the reader to trace any feature back to its original source. Finally, the third appendix contains the full checklist for designing or evaluating an accountability system. The checklist is organized by core concept, and each relevant section is presented as in introduction to a chapter, also organized by core concept.

Literature Review

The literature review includes more than 70 articles, reports, chapters, and full books. The full annotated bibliography can be found in Appendix A. Our initial review of the literature shows one important finding—a surprising lack of conflicting information across the studies. Authors differ primarily in the level of detail provided or in the component of focus. Some authors focused on guiding questions, while others focused on developing a set of accountability standards or core elements. Some authors concentrated on the technical aspects of various indicators while others focused on the goals and consequences of a system. This section summarizes some of the key findings.

Kirst (1990) provides an historical overview of accountability research. He reviewed six broad approaches to accountability, including (1) accountability through performance reporting;

(2) accountability through monitoring and compliance with standards/regulations, (3) accountability through incentive systems; (4) accountability through reliance on the market; (5) accountability through changing the locus of authority or control of schools; and (6) accountability through changing professional roles. Most educational accountability systems include some combination of these approaches. In fact, Raudenbush (2004) argues that accountability systems must include measures of processes, such as information on organizational and instructional practice, in addition to measures of outcomes if the goal is to help improve schools.

Kirst and others argue that a limiting factor in the effectiveness of current accountability systems is erroneous assumptions by policymakers about the nature of schools and their internal capacities. These systems assume that schools have the capacity to improve and lack only the standards, assessment tools, and incentives to do so. However, policymakers must consider what schools will do with the data provided. They then need to develop policies that foster the leadership, collaboration, and skills that are essential to school improvement and to accountability. To strengthen these, schools need both time and resources. Accountability systems can then accommodate and support specific practices and conditions as well as provide the structure of standards, assessment, and consequences.

Much of the literature echoes the same argument and encourages us to look at the system as a whole. An accountability system should be driven by its goals. The indicators of performance selected, the design decisions, and the consequences applied all directly relate to the goals. The goals must be well communicated. The degree of support provided toward meeting the goals and the communication of these goals are strong indicators of the ability of a system to meet its goals. Therefore an overall theory of action should be both explicated and monitored.

Beyond a focus on the goals, authors differed on how they communicated strategies for developing an accountability system. Some created a set of standards or rules to follow, oftentimes grouped into major categories. Other authors highlighted the questions policymakers should ask as they develop an accountability system. Others focused on just one part of the system, such as evaluating the technical quality of the performance indicators or determining effective rewards and sanctions to employ.

Small components of an accountability system have been widely explored. For example, there is a fair amount of research at the school level on the effect of consequences on student achievement. Rather than try to synthesize every study, we chose to focus on the systems on a macro level and examine how the various components work together to build a cohesive system. Many of these specific studies are included in the annotated bibliography for readers who are interested in a specific topic within accountability.

After reviewing this research and various organizational strategies, we created our own organizational strategy as a first step in creating a tool that state or district policymakers could use in developing or evaluating their own system. The next section will describe how we synthesized the current literature and created a matrix to frame our thinking about the various components that comprise all accountability systems.

Synthesizing the Literature

As a first step in synthesizing the information into a useful framework, we created the matrix shown in Appendix B. This matrix allowed us to group common ideas and sort the elements by both major category of an accountability system and the format of the advice given.

The matrix includes core concepts, guiding questions, and key elements taken directly from the literature. Each cell is color coded to indicate the original source. In some cases, an idea came from multiple sources or was developed in discussions with the ASR SCASS, in which case the color was left as black on white.

Core Concepts

We developed a list of seven core concepts based on the various organizational strategies of other authors. Previous researchers and authors differed in how they modeled the essential components. Stecher and Hanser (1992) began with Levin's (1974) definition and elaborated an educational accountability model with four major components, arguing that any effective accountability system must include each of these components:

1. Goals (e.g., 90 percent of high school students will graduate);
2. Measures—means for assessing progress toward the goals (e.g., cost accounting, comprehensive record keeping);
3. A feedback loop—to provide assessment information and constituent input back to the system; and
4. A systemic change mechanism—for reacting to feedback by changing the system, as appropriate.

CRESST (Baker, Linn, Herman, & Koretz, 2002) created a set of 22 standards for educational accountability that they separated into five key categories:

1. System components
2. Testing
3. Stakes
4. Public reporting
5. Evaluation

Hanusheck & Raymond (2002) stated that the "basic skeleton" of accountability systems includes

1. Goals
2. Content standards
3. Measurement
4. Consequences
5. Reporting

According to Carlson (2002), there are five key elements of accountability systems:

1. The goals of the system
2. The selection of key indicators of success and ways to measure them (multiple

measures), rather than merely using information that is available
3. Decisions about how the selected indicators will be scaled, weighted, combined, and reported
4. The types of actions that will be taken based on the resulting performance data (rewards and sanctions)
5. Steps that will be taken to determine and improve the effectiveness of the accountability system itself

And, in their framework for evaluating the validity of an educational accountability system under NCLB, Forte Fast and Hebbler (2004) highlighted four primary components:

1. Performance indicators (Percent proficient, participation rates, graduation/attendance rates, other)
2. Decision rules (how scores are combined and interpreted)
3. Consequences (Sanctions and rewards or changes in PD, pedagogy, or resource allocation)
4. Goals (e.g., 100% proficiency in reading and math by 2014)

Almost all of these organizational schemes include goals and indicators or measures. Several also include consequences or follow-up actions. Reporting, evaluation, and internal monitoring were also mentioned.

To develop our list of core concepts, we started with the four core concepts presented in the validity discussion by Forte Fast & Hebbler (2004). The original category "decision rules" appeared too specific to test scores. Thus, to better meet our purposes we renamed it "design decisions" to focus it on the design of the system and how to ensure the various indicators worked well together. We then added communication and support to emphasize the importance of the reciprocal relationship between state policymakers and local administrators. Finally, work by Carlson, Linn, Baker, and others emphasized the importance of continually evaluating the system itself in order to improve it (Baker, Linn, Herman, & Koretz, 2002; Carlson, 2002). Some authors wrote about building in a feedback loop to provide information and input back into the system and a systemic change mechanism for reacting to the feedback by adapting the system as necessary (e.g., Stecher & Hanser, 1992). We therefore added a final category for evaluating, monitoring and improving the system.

Guiding Questions

The next column in the matrix is called "guiding questions" as it lists questions that are meant to be asked by state (or district) policymakers as they develop or evaluate their accountability system. These questions help to frame the issues for consideration. These questions come from multiple sources but primarily from Ananda & Rabinowitz (2001), Erpenbach (2002), and Gong (2002). The questions generated in these three reports are more alike than different. They all follow the process of designing an accountability system from the establishment of goals, through the design and determination of indicators to the effects and evaluation.

Ananda & Rabinowitz (2001) identified a sequence of key questions that must be addressed in planning an accountability system. They focused on the big-picture issues in designing

an accountability system asking question starting from the first issue in accountability system that is "What are the primary goals you are trying to accomplish with an accountability system?" through the question of what to do with the results "What will you do about the problems uncovered through the accountability system?" They laid out a simple sequence of questions and issues from the goals to the indicators to the consequences to the effects.

Erpenbach (2002) and colleagues focused primarily on the use of multiple measures in developing accountability systems. They developed ten questions moving from the purposes of the system to the selection of multiple indicators, to decisions on how to combine the indicators, to reporting, to finally evaluating the effects and the system itself.

Gong (2004) developed an earlier report with the ASR SCASS creating a linear sequence of ten questions that state policymakers could use when designing an accountability system. The questions were developed to include criteria and comments intended to "provide a structure for helping states moved through the process of designing a school accountability system." (p. 2). The questions start with defining the purposes, move to defining how the data will be collected, combined, and reported and conclude with how the system itself will be evaluated.

Again, all three have common elements reflected in the guiding questions section of the report. They all include the same sequence of events. They all highlight the importance of stating clearly the goals and purposes of the systems. They all discuss the performance indicators and design decisions. They all include a consideration of an internal evaluation and monitoring system.

Key Elements

The third column, Key Elements, include both standards that researchers and practitioners recommend for inclusion in any accountability system and considerations that are more dependent on the goal.

The most commonly cited standards include those developed by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Baker, Linn, Herman, & Koretz (2002) developed a set of 22 standards for accountability systems in five categories: system components, testing, stakes, public reporting, and evaluation. These standards represent models of practice from three perspectives: research knowledge, practical experience, and ethical considerations.

Hill and DePascale (2003) provide a framework for evaluating technical considerations of the accountability system. The standards developed in their work involve tradeoffs between validity and reliability and inform such issues as minimum group size and confidence intervals.

Other key elements come from Elmore (2003) who developed a set of principles after studying two low-performing schools and the reforms they implemented. His standards focus most on supporting schools as they work to improve student outcomes. Elmore (2004) continues to focus on the reciprocity between accountability and capacity, emphasizing the importance of setting realistic goals. Forte Fast and Hebbler (2004) focus on consistency within the entire system. That is, once an explicit theory of action has been developed, care should be taken to match indicators and decision rules to the goals. The consequences should

also be implemented to support the realization of the goals. Hamilton and Stecher (2004) and Fuhrman (1999), among others, provide some guidance on the use of incentives and sanctions.

Options/Examples

The final column in the matrix includes various options or examples. These include different ways to approach the key elements. Some of the options or examples are very specific to a particular goal, while others offer advice that would work with almost any system. For example, a key element is to include multiple measures as performance indicators, under options/examples, some suggestions are given such as teacher qualifications, curriculum, and learning environments.

Another example of an option included in the last column is to determine the appropriate number and level of grades to include in the system. Prior to NCLB, many state accountability systems focused on "key" grades, such as grades 4, 7, and 10. The federal requirement is to include all grades 3–8 plus one high school grade in the accountability system. For a system not focused on NCLB requirements, determining which grades to include should be a key consideration.

Furhman (1999) gives examples of the tradeoffs between complexity and fairness. For example, an index system may provide the most balanced measure of a school's performance, but it can be difficult to explain to school administrators and the public.

Darling-Hammond (2006) provides examples of multiple measures to include in an accountability system, as well as guidance on how to develop assessments, administration guidelines, and performance reports.

Linn (2005a) addresses the options of using compensatory or conjunctive accountability systems. Even within a conjunctive system, he addresses the options of having all subgroups meet the same goal compared to setting individual goals for each subgroup depending on their starting point. He also discusses the option of letting the annual performance goal determine the end point versus setting a fixed end point for all students.

Several authors provide options and examples for supporting districts and schools as they work to achieve the accountability goals. For example, Ananda and Rabinowitz (2001) discuss the importance of supporting schools through technical assistance and financial support. The American Institutes for Research (AIR) has investigated support systems at the district level and focuses on five main areas of support (AIR, 2006):

1. Needs assessment and planning (revisions of district plans, data analysis tools to identify where to target resources)
2. Data analysis (state provides the analysis or tools to help districts manage large data sets)
3. Capacity building (focusing on enhancing districts' PD, leadership, or curricular activities)
4. Resource allocation (aligning budgets with priorities, providing tools to assist districts in managing and targeting funds)

5. Progress monitoring (conducting district audits and onsite reviews)

Elmore (2003) focuses on supports teachers need to help improve student learning. These supports are in the form of financial, including loan forgiveness and extra compensation; professional development, such as access to courses and advanced degree program); and improved working conditions, including reduced teaching loads, improved physical infrastructure, and better materials for the classroom.

Ohio Department of Education has developed a strategic plan that includes supporting the accountability goals by communicating clear, high expectations to students; aligning the teachers' professional development to the district's mission; providing teachers sufficient time for planning; providing students with appropriate opportunities to learn; and implementing appropriate instructional interventions.

Further and more detailed examples of state strategies for meeting their goals within each core concept are provided in Chapter 3. The above examples are intended to be brief mentions of the types of options that may be used for different components of an accountability system. Specific examples of options states have used for each of the core concepts are provided in subsequent chapters.

Differences between Federal and State Accountability Systems

One important theme in the literature since the passage of NCLB is the challenge of incorporating existing state accountability systems into the federal accountability requirements. Though the *Improving America's Schools Act of 1994* (IASA), a reauthorization of the *Elementary and Secondary Education Act*, helped to kick off the standards movement, implementation of statewide standards was inconsistent. For example, although IASA required states to develop a single statewide accountability system, many states were not yet in compliance with this act and had varying degrees of state level accountability in place in 2001. NCLB brought more pressure on the states to develop strong accountability systems that met very specific criteria; however, it did not require one single accountability system. A survey by CCSSO found that most states adjusted their own accountability systems to meet the federal requirements and created one system to serve both purposes. Other states did not have a well-developed state accountability system at the time NCLB was passed and developed one to meet the new federal requirements. A handful of states, however, have attempted to maintain their own state accountability system and adopted a separate system to meet the federal requirements.

For purposes of simplification, we will call the federal accountability system the *Adequate Yearly Progress* system or AYP. Consider a state such as Florida that has maintained its A+ accountability system or California that still uses the API to rate schools. These states may produce different classifications for their schools when using their state system and AYP.

Linn (2005b) has written about the mixed messages schools receive from such bifurcated systems. That is, a school might meet the goal under the state accountability system but not meet it under the federal system, or vice versa. For example, Kentucky uses an index based on tests in seven content areas, attendance, and graduation rates. Targets are set biennially by

school. In 2004, 74% of Kentucky schools made federal AYP targets, while 95.6% met their state targets. Not all schools who met AYP met state targets. Approximately 25% of schools met state targets while missing AYP, providing mixed messages to teachers, administrators, and the public.

In Florida, 56% of schools that received an A on the state accountability system in 2004 failed to make AYP. In Colorado in 2003, 21.9% of schools rated as "unsatisfactory" and 47.5% of schools rated as "low" made AYP, while 13.7% of schools rated as "high" failed to make AYP.

Linn (ibid.) discusses several features of NCLB that make it likely to have results conflicting with state accountability system results:

- The use of absolute targets
- The need for conjunctive targets
- The requirement of meeting targets by subgroup

Every school has between 5 and 37 targets (hurdles) under NCLB, meaning there are many ways to miss AYP and only one way to make it. That is, a school must meet its target for every single subgroup to make AYP, but if it misses just one – no matter which one or how many over one – it is placed into an improvement category. Novak and Fuller (2003) found that schools serving more diverse students were less likely to meet AYP requirements than schools serving less diverse student bodies.

Other differences in ways in which school achievement benchmarks are set for NCLB compared to state accountability systems include:

- NCLB focuses on status measures, while state systems typically incorporate improvement or growth measures
- NCLB sets an absolute level of performance regardless of school's starting point, whereas state systems typically set goals based on the school's baseline performance.
- NCLB and states differ on the long-range goals and the timeline for meeting them.

Goertz (2001) also identifies how state accountability systems varied prior to the passage of NCLB. Her analyses correspond well with those by Linn. She discusses how state systems tended to use a progress goal rather than an absolute goal. Thus, the expected level of student performance was different then than under NCLB. Other variations between state systems and NCLB included the percentage of students that schools were required to have meet this expected level and the length of time schools were given to meet their goal. All of these factors are now regulated under NCLB.

States that maintained their previous accountability system or who later supplemented the AYP model with their own state model now face a situation in which school may be shown to be high quality under one system and poor performing under another. As will be discussed in the next chapter, policymakers need to consider how these mixed messages may affect school reform efforts and parent decision making.

Summary

The literature on this topic is extensive. In fact, there is much more literature than what has been summarized in this chapter. The next chapter continues the literature review in greater detail, but in the context of the core concepts listed in the matrix. Overall, it is worth noting the extent to which different researchers and educators agree on the necessary components of an accountability system.

The matrix that synthesized the literature review was reformatted into a checklist tool, assembling key elements in a manner that can be used by states. There are seven core concepts as described in the previous chapter. Each of these concepts is used as an organizer. Within each of the core concepts, we have listed guiding questions and key elements. The guiding questions are intended to help policymakers frame their system as they explore their answer to each question. The key elements include both standards of what should be done and considerations of what can be done. For instance, "develop an explicit theory of action" is a standard that should be followed, while "consider whether to use all grades or selected grades" is not a directive, merely a reminder that this is an issue that should be deliberated. There are areas where (1) we do not know the most effective model, such as the best type of incentive to use, or (2) the most effective model is completely dependent on the goal of the system. Since we do not presume to know every state's goal(s), we cannot provide exact guidance on the best option to select for each element. The purpose of this tool is not to endorse one model over another but to ensure that those using it consider all possible options and use best practices when they are known and appropriate.

The organizing structure focuses on seven core components that must be considered in any educational accountability system:

1. Goals
2. Performance indicators
3. Design decisions
4. Consequences
5. Communication
6. Support
7. System evaluation, monitoring, and improvement

Each of these components is addressed in a subsequent chapter.

References and Suggested Readings for This Chapter

American Evaluation Association (2002). *Public statement: High stakes testing in preK-12 education*. Fairhaven: MA: American Evaluation Association. Downloaded October 8, 2007 from http://www.eval.org/hst3.htm.

American Institutes for Research. (2006). *Summary of state strategies for districts identified for improvement under NCLB*. Sacramento, CA: California Comprehensive Center.

American Psychological Association, American Educational Research Association, & the National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Ananda, S. & Rabinowitz, S. (2001). *Building a workable accountability system*. San Francisco: WestEd. Downloaded February 20, 2007 from www.WestEd.org/online_pubs/kn-01-03.pdf.

Bailey, A., Winter, P., & CAS SCASS. (2002). *Incorporating federal requirements into state accountability plans*. Washington, DC: Council of Chief State School Officers.

Baker, E., Linn, R., Herman, J., & Koretz, D. (2002). *Standards for educational accountability systems* [Policy Brief 5]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing

Carlson, D. (2002). The focus of state educational accountability systems: Four methods of judging school quality and progress. In W. J. Erpenbach et al., *Incorporating multiple measures of student performance into state accountability systems—A compendium of resources* (pp. 285-297). Washington, DC: Council of Chief State School Officers.

Darling-Hammond, L. (2006). *Standards, assessments, and educational policy: In pursuit of genuine accountability*. Princeton, NJ: ETS

Elmore, R. (2003). *Knowing the Right Thing to Do: School Improvement and Performance Based Accountability*, Washington, DC: NGA Center for Best Practices

Elmore, R. (2004). Conclusion: The Problem of Stakes in Performance-Based Accountability Systems. In Fuhrman, S. & Elmore, R., (2004). *Redesigning Accountability Systems for Education*. New York, NY: Teachers College Press.

Erpenbach, W. (2002). *Incorporating multiple measures of student performance into state accountability systems—A compendium of resources*. Washington, DC: Council of Chief State School Officers.

Forte Fast, E. & Hebbler, S. (2004). *A framework for examining validity in state accountability systems*. Washington, DC: Council of Chief State School Officers

Fuhrman, S. (1999, January). The new accountability. *CPRE Policy Briefs*, RB-27. Philadelphia, PA: The Consortium for Policy Research in Education.

Fuhrman, S. & Elmore, R., (2004). *Redesigning Accountability Systems for Education*. New York, NY: Teachers College Press.

Goertz, Margaret. (2001). The Federal Role in an Era of Standards-Based Reform. in *The Future of the Federal Role in Elementary and Secondary Education: A Collection of Papers*. Washington, DC: Center for Education Policy.

Gong, B. & ASR SCASS. (2002). *Designing school accountability systems: Toward a framework and process*. Washington, DC: Council of Chief State School Officers.

Hanushek, E. & Raymond, M. (2002). Sorting out accountability systems. In W. Evers & H. Walberg (Eds.) *School accountability* (pp. 75-104). Palo Alto, CA: Stanford University, Hoover Press.

Hill, R. & DePascale, C. (2003). Reliability of No Child Left Behind accountability designs. *Educational Measurement: Issues and Practice*, 22(3), 12-20.

Joint Committee on Standards for Education Evaluation. (2004). *The Program Evaluation Standards. How To Assess Evaluations of Educational Programs*. (2nd Ed.). Thousand Oaks, CA: Sage Publications.

Joint Committee on Standards for Education Evaluation. (2002). *The Student Evaluation Standards. How To Improve Evaluations of Students*. Thousand Oaks, CA: Corwin Press.

Kirst, Michael W. (1990). *Accountability: Implications for State and Local Policymakers*. Washington, DC: Information Services, Office of Educational Research and Improvement, U.S. Department of Education.

Levin, H. (1974). A Conceptual Framework for Accountability in Education. *School Review* 82(3), pp. 363-391.

Linn, R. (2005a). *Test-based educational accountability in the era of No Child Left Behind*. [Technical Report 651]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Linn, R. (2005b). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *Educational Policy Analysis Archives*, 13(33). Retrieved January 15, 2007 from http://epaa.asu.edu/epaa/v13n33.

Novak, J. R. & Fuller, B. (2003, December). Penalizing diverse schools? Similar test scores, but different students, bring federal sanctions. *Policy Brief* 03–4, Berkeley, CA, University of California Berkeley: Policy Analysis for California Education, available online at http://pace.berkeley.edu.

Raudenbush, R. (2004). *Schooling, statistics, and poverty: Can we measure school improvement*? ETS Angoff Lecture, 2004.

Stecher, B. & Hanser, L. (1992). Local accountability in vocational education: A theoretical model and its limitations in practice. MDS-291. Santa Monica, CA: RAND.

Chapter 2: Goals of Accountability Systems

Checklist for State Accountability Systems: Goals

- *What are the purposes of the accountability system? Why are you implementing the system?*
  - □ Explicate the local, state, and federal requirements that this system is trying to fulfill.
- *What are the primary goals you are trying to accomplish with an accountability system? What are you hoping to accomplish?*
  - □ Establish and prioritize goals.
- *What accountability decisions will be made and with what consequences?*
  - □ Develop an explicit theory of action that relates the goals and design of the system to anticipated outcomes by the state, schools, and others.
- *In what systems are you working? What are the main legal and policy constraints or specifications?*
  - □ Explore the legislative, policy, and legal influences and understand how those considerations will determine the design and implementation of the accountability system.
- *In broad terms, what are the schools/students (or others) to be held accountable for?*
  - □ Match the indicators and decision rules to the goals.
  - □ Begin to consider how to combine multiple indicators to match the goals by determining whether the goals best fit a status, improvement, or growth model.

The goals portion of the framework emphasizes the purposes, uses, and contexts for the accountability system. We distinguish between the purposes, which provide an overarching reason for using an accountability system, and the goals, which specify the intended outcomes. Third, we introduce the idea of decisions and consequences here, although they each are considered a core concept and have a separate section later. It is important to start thinking about decisions and consequences when developing a theory of action. Developing a theory of action is one of the key elements stressed here as it links the intended outcomes to the various indicators and supports provided. The next key element focuses on understanding the context in which the system must operate and the external pressures that may constrain certain measures or decisions. For example, legal restrictions on individually identifiable data may require a state to set a minimum number of students that must be in any category before data can be reported in that category. Finally, we must begin to think about the system in broad terms under this first core concept, focusing on the relationship between the goals, intended outcomes, measures, decisions rules, supports, and consequences.

Linn (2001) encourages state policymakers to be clear about the intended purpose(s) for their educational accountability system. For example, while most states or districts would agree

that the purpose of accountability is to improve student learning, Linn argues that states need to be more specific in stating their priorities for achieving such a goal. For instance, beyond improving student learning, state policymakers may specifically desire to:

- Reinforce content standards in priority subjects
- Support deep understanding and problem-solving ability
- Assure a given level of achievement for students before promotion

He also recommends that policymakers clarify the uses of their system, asking specific questions such as:

- What results will be given to parents?
- What will be done with school-level results?
- How much emphasis should be placed on status versus improvement?

The report by Gong and the ASR SCASS (2002) focused on the importance of understanding the legal and policy constraints surrounding the system. Policy contexts, in particular, not only drive the implementation of an accountability system, but may change as an administration changes requiring constant updating of a system. Erpenbach (2002) also considers legislative and policy influences a crucial consideration in developing goals and designing the accountability system. On one level, NCLB has driven the policy agenda to focus on a basic skills model with its requirement of 100% proficiency. By requiring 100% of students to reach the bar, the legislation has effective negated the importance of achievement beyond that bar. And, as shifts are made in the law – allowing a growth model, for example – accountability systems shift to adjust for these changes.

Other legal considerations may focus more on the specifics within a system. Legal requirements are particularly important when considering issues of access and equity. For example, Baker, et al. (2002) highlighted one legal constraint in their comments on Standard 4 in saying "Legal requirements as well as ethical considerations demand that all students be included in the accountability system." Whether these types of policy and legal influences are positive or negative can be debated, but it is important to acknowledge their existence and their ongoing effect on the accountability system.

Ananda and Rabinowitz (2001) discuss the problem many states have with trying to set too many goals. For instance, they discuss the pressures policymakers feel to "embrace multiple sweeping education improvement goals, such as improving student learning, motivating teachers and students, reducing achievement disparity between majority and minority students, monitoring education costs, improving access to education, building public confidence in education, and improving the state's competitive economic status as compared to other states." (p. 2). They recommend state policymakers consider selecting short-term and long-term goals or phasing in the different goals. There is a widely held belief that systems with too many goals will fail to achieve any. It is important to focus the goals to make them attainable.

Theory of Action

Several researchers and state policymakers emphasized the importance of developing a coherent theory of action that relates the goals, accountability system design, consequences, communication, and support to anticipated outcomes by the state, schools, and others. The theory of action explicates the policymakers' assumptions about how the accountability system will bring about the desired changes. This step is necessary to judge the validity of a system and determine whether or not it is effective and why.

> "This theory of action governs the system's goals and objectives and is presented so as to promote a deeper understanding of Delaware's reform initiatives."

The standards-assessment-and-accountability-based educational reforms, of which NCLB is an example, share a general theory of action that might be stated:

By

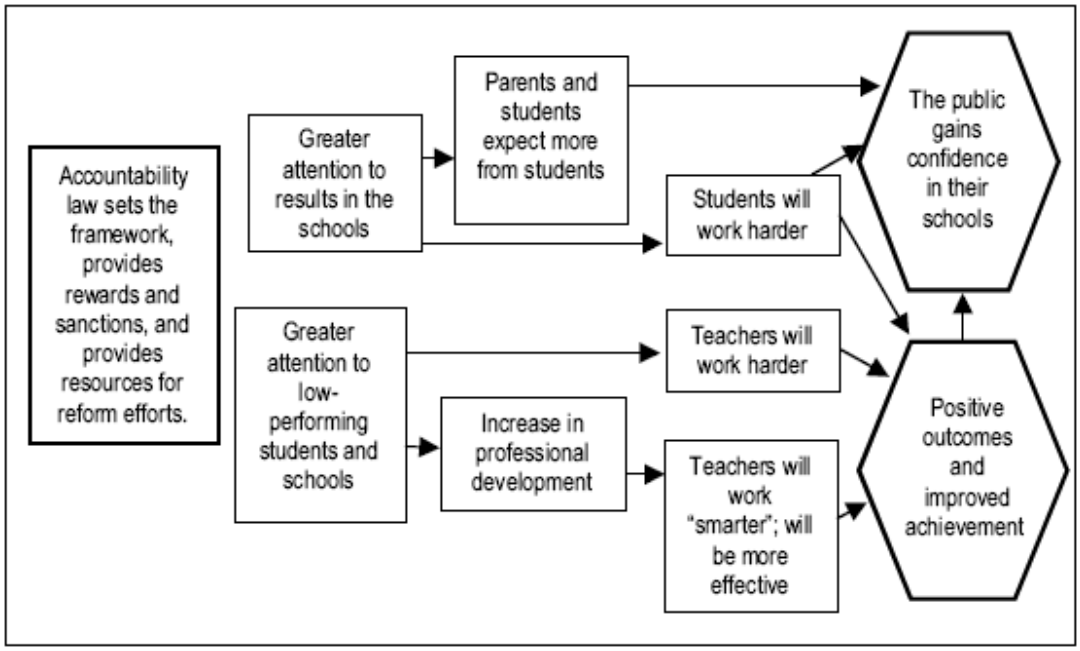
- Clearly stating the goals (100% of students proficient by 2013-14, and closing any achievement gaps of subgroups), and
- Measuring performance, primarily through valid assessments aligned to important content knowledge and skills, and
- Providing additional financial and other resources, and
- Holding states accountable by reporting the results publicly and providing progressive assistance and sanctions to those schools that do not perform adequately,

the following actions will result:

- States will be motivated to provide instructional services to all students in order to meet the goals. Where students in schools do not perform adequately, schools will devise effective means to help the students/subgroups improve their learning.
- When schools are not able to improve student performance adequately or when a school does not perform adequately according to AYP requirements, students will be able to transfer to a public school that will provide them better learning opportunities. Schools identified as in need of improvement will receive appropriate assistance and graduated sanctions that will help the schools educate their students adequately as defined by AYP.
- Clear public reporting will provide schools the information they need to identify needs accurately, and will enable parents and others to see where students and schools need to improve and bring appropriate pressure to bear.

Marion et al. (2002) provide the following figure of an "over-simplified" version of how various elements in an accountability system relate to each other. They state that "visual displays allow all the assumptions and steps to be considered simultaneously, so the missing ones might be spotted. More importantly, the very process of agreeing on the theory will encourage essential discussion among the policymakers and others responsible for the design of the accountability system." (p. 35).

Figure 1. A Simplified Theory Of Action: How Accountability Expectations And Actions Will Lead To Higher Achievement



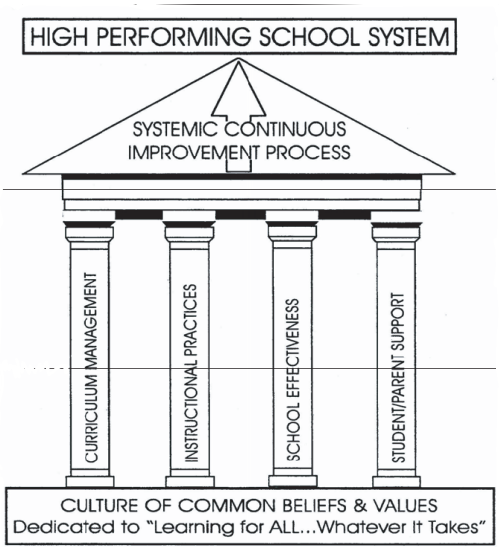
Taken from Marion, S., White, C., Carlson, D., Erpenbach, W, Rabinowitz, S., & Sheinker, J. (2002). *Making valid and reliable decisions in determining adequate yearly progress.* Washington, DC: Council of Chief State School Officers. p. 36.

Several state policymakers have taken this idea and developed an explicit theory of action for their state that depicts how its actions and resources will be used to achieve its goals both within and beyond NCLB. For example, Nebraska's operational theory of action has three main goals: improving student achievement, maintaining efforts in school improvement, and building the capacity of educators to implement needed changes. As described by the commissioner of education, Doug Christensen:

> Nebraska's approach to standards, assessments, and accountability—School-based Teacher-led Assessment and Reporting System (STARS)—is firmly grounded in the belief that decisions about student learning should be standards-based and should be focused on classroom knowledge of the student. This process relies upon the professional expertise of Nebraska educators and has been built upon a statewide initiative to develop educator capacity in assessment design and the use of assessment data for improved instruction. The requirements of federal NCLB have been integrated into the accountability requirements.

Nebraska's theory of action and thus its system differs from that of most other states in that it is locally based. Teacher-designed classroom assessments are the foundation of the system rather than statewide assessments. The state role is to support the development of classroom and district assessments through specialized training and professional development activities. The values embodied by this approach include providing real-time information back to teachers and embedding improvement activities into everyday practice of school and district staff.

West Virginia symbolizes its theory of action in a graphic that illustrates the major components of the Framework for High Performing School Systems. Based on an analysis of high performing schools, policymakers in West Virginia believe that high performing school systems share three broad commonalities: (1) Strong core beliefs that shape the culture of the system; (2) System-wide strategies that enhance curriculum, instruction, overall school effectiveness and student/parent support; and (3) Use of a systemic continuous improvement process to bring about change.



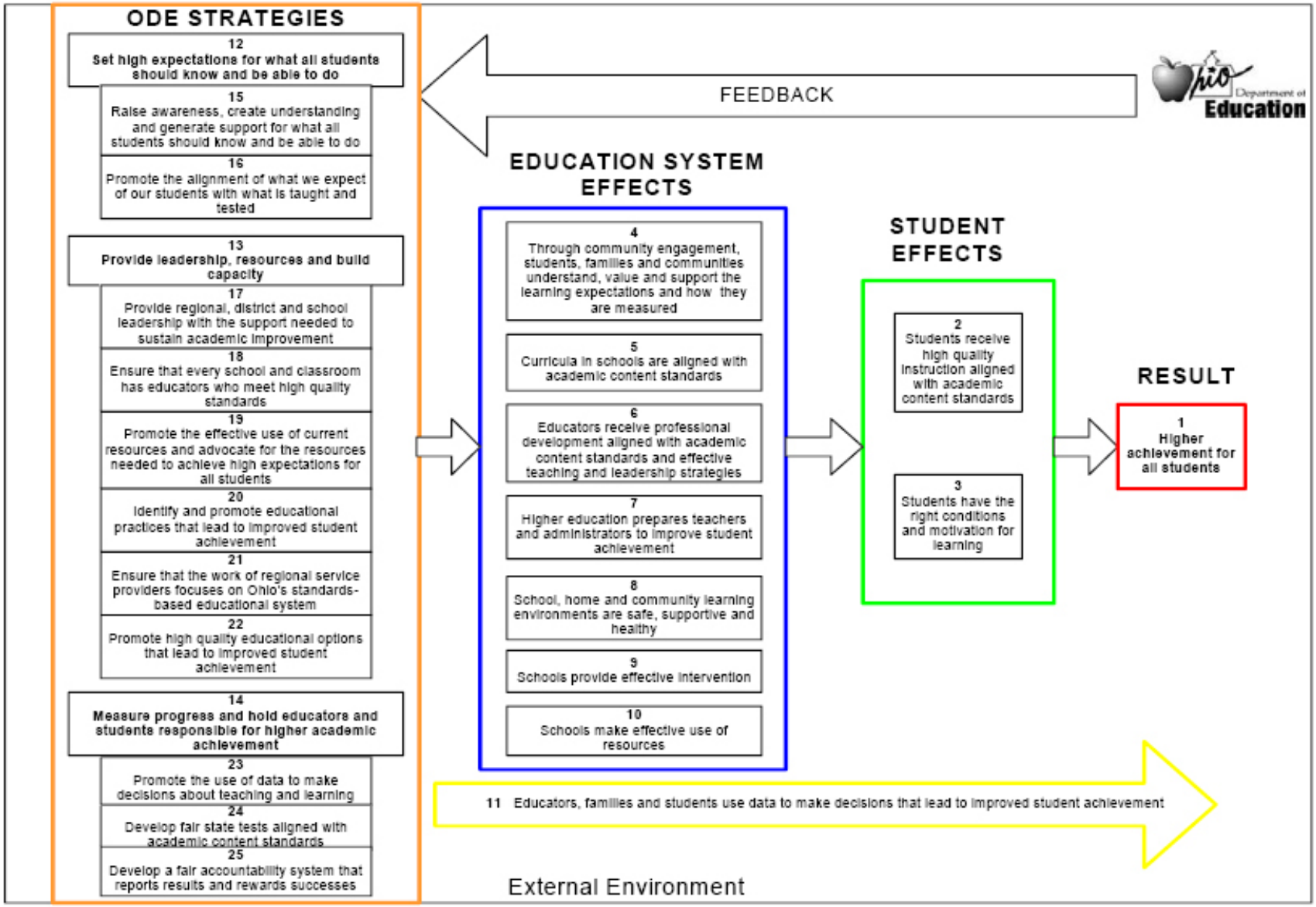
As another example, the Ohio Department of Education (ODE) has developed a strategic plan that lists three missions and shows how department strategies and success indicators will lead to the desired results of higher achievement for all students. ODE considers its theory of action to be a living document in the sense that as new information is gathered and analyzed ODE may revise some connections, add or modify their indicators, and add new strategies. This information reflects their strategic plan prior to receiving results from the 2006–07 school year. As seen on the following page, ODE has stated its result or vision to be "higher achievement for all students" and implemented strategies to achieve this result. In between the strategies and results is a series of success indicators subdivided into education system effects and student effects. Ohio has more extensive documentation than most states, and uses this explicit theory of action to organize its efforts internally, allocate resources, and direct program evaluation efforts.

ODE has developed three primary strategies—set high expectations for what all students should know and be able to do; provide leadership, resources and build capacity; and measure progress and hold educators and students responsible for higher academic achievement—and listed two to six sub-strategies for each. Each sub-strategy also has a list of measures that the state can use for monitoring its success. (See the following page for the details on the measures.) For instance, under the sub-strategy *Raise awareness, create understanding, and generate support for what all students should know and be able to do*, ODE has developed four measures:

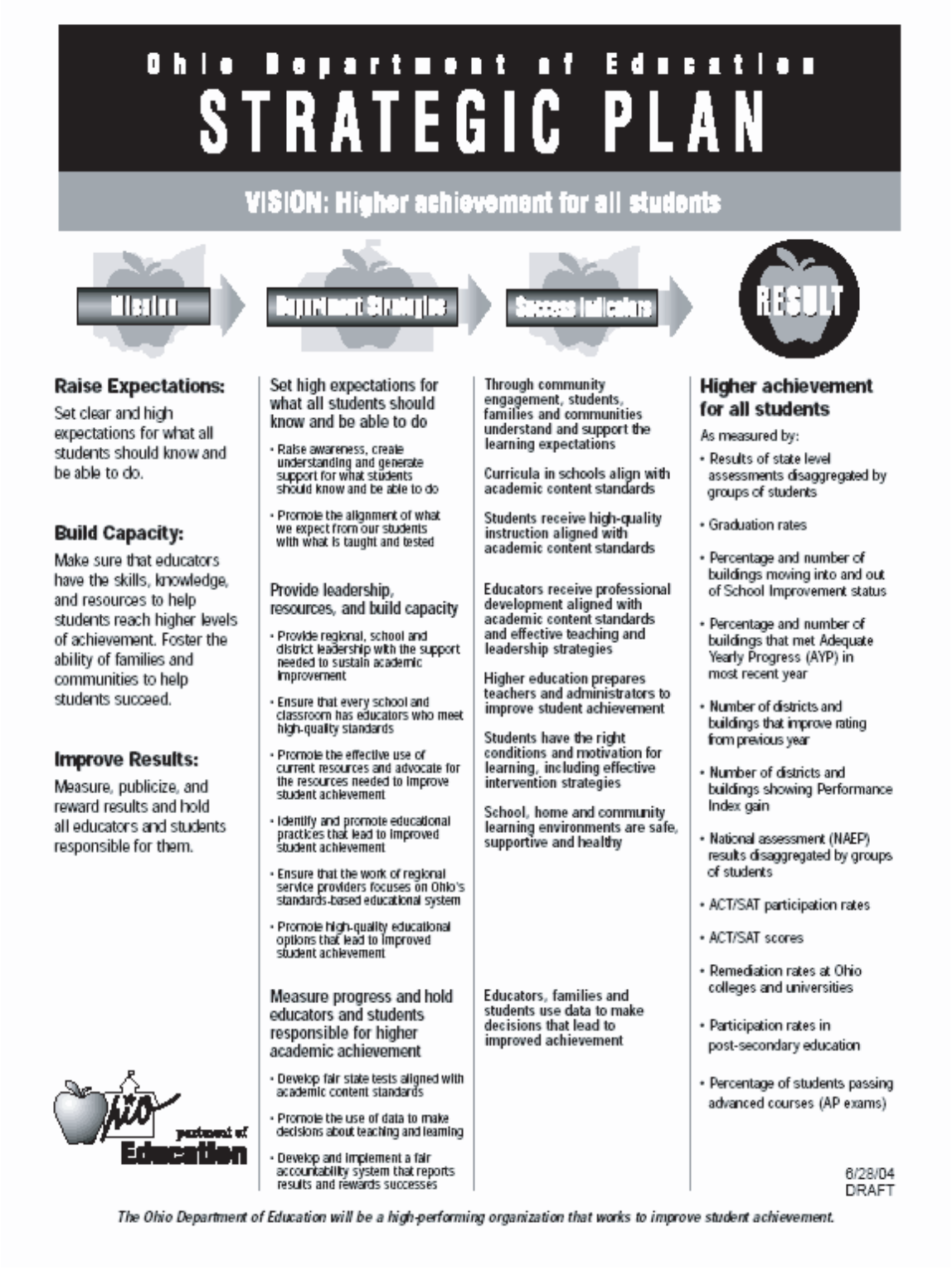
1. Individual student report cards, developed and distributed by school year 2005–2006
2. Percentage of target audiences receiving publications on student learning expectations
3. Number of parents/families attending state-sponsored information sessions
4. Percentage of counties with community-based organizations partnering with ODE.

ODE has also espoused a theory of action for the effects implementing these strategies would have on both the education system and students. They call these effects success indicators and

Ohio Department of Education (ODE) Theory of Action



describe them as elements that must be in place to achieve their desired result. Each indicator also has a series of measures associated with it to monitor the degree to which these effects are occurring. For instance, the success indicator *Educators receive professional development aligned with academic content standards and effective teaching and leadership strategies* has three measures associated with it:



1. Percentage of teachers participating in "high-quality professional development"
2. Percentage of ODE-sponsored professional development activities that meet the standards of "high-quality professional development"
3. Number of teachers and administrators completing Entry Year Programs

Finally, ODE lists 12 measures for the results of "higher achievement for all students." In addition to the federally mandated measures of percentage of students, schools, and districts meeting AYP, this system also looks at measures of improvement (number of districts and buildings that improve rating from previous year), growth (number of districts and buildings showing Performance Index gain), and other measures such as graduation rates, participation in post-secondary institutions, percentage of students passing advanced courses (AP exams). The system also considers performance on non-state tests, such as NAEP, ACT, and SAT.

ODE states their belief that this aligned education system will promote higher achievement for all students. By constantly monitoring success at all stages of the system, they can evaluate and adjust their strategies to ensure they are implemented effectively. Furthermore, this type of documentation addresses at least three of the four aspects of a theory of action. The one piece that is not clearly articulated is their theory for the consequences. One sub-strategy indicates that developing a fair accountability system includes rewarding success, and we know that specific sanctions are required for schools and districts that miss AYP two years in a row, but this theory of action does not explain how the rewards or sanctions will lead to improved education system or student effects and thus work to meet the desired result.

References and Suggested Readings for This Chapter

Ananda, S. & Rabinowitz, S. (2001). *Building a workable accountability system*. San Francisco: WestEd. Downloaded February 20, 2007 from www.WestEd.org/online_pubs/kn-01-03.pdf.

Baker, E., Linn, R., Herman, J., & Koretz, D. (2002). *Standards for educational accountability systems* [Policy Brief 5]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing

Erpenbach, W. (2002). *Incorporating multiple measures of student performance into state accountability systems—A compendium of resources*. Washington, DC: Council of Chief State School Officers.

Forte Fast, E. & Hebbler, S. (2004). *A framework for examining validity in state accountability systems*. Washington, DC: Council of Chief State School Officers

Gong, B. & ASR SCASS. (2002). *Designing school accountability systems: Toward a framework and process*. Washington, DC: Council of Chief State School Officers.

Hamilton, L., Stecher, B., & Klein, S. (Editors.) (2002). *Making sense of test-based accountability in education*. Santa Monica, CA: RAND

Linn, R. (2001). The design and evaluation of educational assessment and accountability systems. [Technical Report 539]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Marion, S., White, C., Carlson, D., Erpenbach, W, Rabinowitz, S., & Sheinker, J. (2002). Making valid and reliable decisions in determining adequate yearly progress. Washington, DC: Council of Chief State School Officers.

Checklist for State Accountability Systems: Performance Indicators

- *What data are available and will be used in the accountability system?*
  - □ Employ different types of data from multiple sources.
  - □ Think broadly about types of measures to include, such as measures of teacher qualifications, curriculum, and learning environments.
  - □ Include data elements that allow for interpretations of student, institution, and administrative performance as appropriate.
  - □ Avoid making decisions about individual students, educators, or schools on the basis of a single indicator.

- *Which students should be included in the system?*
  - □ Include the performance of all students, including subgroups that historically have been difficult to assess.
  - □ Determine how to include special populations.
  - □ Consider whether to use all grades or selected grades.
  - □ Determine how to include alternative education programs (such as vocational schools or schools intended for short-term attendance by students with disciplinary problems).
  - □ Determine whether to include students new to the school or district.

- *When should indicators be collected?*
  - □ Consider when you need the data to make accountability decisions.
  - □ Decide whether your system's goals support fall testing, spring testing, or more frequent testing.
  - □ Eliminate artificial barriers to students demonstrating ability, such as time limits and inflexible administration.

- *How do you ensure the data for the indicators are interpreted correctly and/or produce valid results?*
  - □ The accountability system is only as strong as its indicators.
  - □ The validity of the measures that have been administered as part of an accountability system should be documented for the various purposes of the system.
  - □ Develop a quality assurance plan that examines the accuracy and validity of all indicators.

- *What technical issues and additional analyses will need to be addressed in developing a valid set of indicators?*
  - □ If tests are to help improve system performance, document that test results are modifiable by quality instruction and student effort.
  - □ Performance components should provide appropriate representation of the knowledge and skills valued by the state and/or district.
  - □ Reliability of other indicators used as outcome measures, such as teacher qualifications and graduation rates, should be documented.

A major issue in any accountability system is the question of what to measure. To answer that question, we need to examine the data that are available, the targets of the data collection, and the timing of the data collection. Once those pieces have been decided, consideration also needs to be given to ensuring the reliability and validity of the data used to inform the accountability system.

Researchers strongly suggest that systems include multiple measures. No student, school, or district should be sanctioned on the basis of one indicator, such as a test. According to Raudenbush (2004), "Accountability systems must include measures of processes, such as information on organizational and instructional practice, in addition to measures of outcomes if the goal is to help improve schools." Such measures may include teacher qualifications, curriculum, learning environments, attendance rates, graduation rates, school size, class size, responsiveness to parents, and teacher preparatory time. For research on education indicators, also see Shavelson, et al, RAND, ; Blank, 1993; and the CCSSO annual reports on state education indicators,, http://www.ccsso.org/projects/State_Education_Indicators.

Porter (1991) recommends a list of indicators for measuring school quality including:

- Instructional leadership
- Existence of a focused mission
- High expectations for all students
- Task-oriented climate
- System for monitoring outcomes within the school
- School and class size
- Grade level organization
- Teacher planning time
- Availability, quantity, and credentials of teacher aides
- Number of teacher preparations required
- Shared decision making

Porter also recommends gathering data at the state or district level on the following elements:

- Subject matter framework
- Student assessments
- Control of textbooks
- Specifying graduation requirements
- Guidelines for allocating time to core subjects
- School accreditation programs
- Management information systems

Even in focusing on student performance as measured by assessments, policymakers need to consider which subjects will be included in the accountability model. NCLB bases decisions solely on reading and mathematics performance, but several state systems also include science, social studies, and writing.

For example, California's accountability system includes both their federal AYP goals and indicators, but they use their state index—the Academic Performance Index (API)—as their additional indicator. The API is a weighted index that includes results from both their crite-

rion-referenced and norm-referenced assessments in ELA, math, science, and history/social sciences—different indicators are used at different grade levels. More information is provided on the API in chapter 4, but it is important to note that California has made the policy decision to include science and history/social sciences in their accountability system, even though those subjects are not required to be used to determine a school's classification under the federal system.

Kentucky also uses results from their science and social studies tests in their state accountability system. In addition, they include a writing portfolio at grades 4, 7, and 12, test arts & humanities at grades 5, 8, and 11, and practical living & vocational studies at grades 5, 8, and 10.

In addition, while NCLB uses all grades 3–8, policymakers should deliberately consider whether assessing every grade makes sense, or if they want to measure progress at key grades, perhaps one at each school level: elementary, middle, and high. Prior to NCLB, many states had assessments at "key" grades, such as grades 4, 7, and 10. If the requirement to test at every grade 3–8 did not exist, what testing schedule would make the most sense in achieving the goals of the system? Likewise, if policymakers decide to include additional subjects beyond reading and math in their accountability system, how often should achievement in the other subjects be assessed?

Additional Indicators

Missouri uses the following indicators in their state accountability system in addition to the traditional achievement and participation measures:

- ACT (percent of graduates scoring at or above the national mean)
- Advanced Courses (percent of credits earned)
- Career Education Courses (percent of credits earned)
- College Placement (percent of graduates entering college)
- Career Education Placement (percent of career education completers placed in occupation)
- Graduation & Attendance Rates

NCLB requires one additional indicator to supplement the participation and proficiency indicators. Most states have chosen to use attendance rates at the elementary and middle school levels and graduation rates at the high school levels. However, some states have gone a step further by including at least one other additional indicator to those required by NCLB. For example, Delaware uses a vertical scale across grades 3 through 8 to report their assessment results. An additional indicator in their accountability system, then, is whether students are maintaining or showing progress from one grade to the next. Delaware, Kentucky, and California—as well as other states—also consider student performance in science and social studies as part of an additional indicator. Oklahoma and Ohio use additional indicators in their high schools that measure the percentage of students receiving advanced placement (AP) credit and taking the ACT or SAT.

Another consideration besides simply adding more indicators is to differentiate the indicators based on school or district characteristics. That is, consider using different indicators of excellence in rural versus urban schools or in schools with high diversity. Some states already use

differentiated indicators across different school levels (e.g., elementary versus high school). As stated in the example above, Oklahoma includes information on academic excellence at the high school level that does not exist at the elementary or middle school level, including performance on the ACT and AP tests and the need for college remediation. These indicators are part of their academic index at the high school level, but there are no equivalent indicators at the elementary or middle school level.

Validity of the Indicators

The validity for each measure should be documented. Because a major indicator in any accountability system is performance on an assessment, the use of the assessments should be valid. There are many standards that address the validity of test interpretation in the AERA/APA/NCME *Standards for Educational and Psychological Testing* that should be followed. Some areas that help build the validity argument include demonstrating the alignment of the test items with the content standards, using a representative sample of judges to set cut scores, providing the rationale for creating subscores or composite scores, and including interpretation aids in score reports.

Validity for assessment is generally considered to have two aspects: The assessment measures what it intended to measure, and the interpretations and uses (consequences) are justified. Similarly, validity for accountability decisions should reflect accurate evaluation of what was intended to be measured and justification of the interpretations and uses (especially consequences formally specified as part of the accountability system). By extension, some theorists would hold that an accountability system is fully valid only if its consequences are also fully valid.

Although the validity of an accountability system builds on the validity of assessment systems, having a valid assessment system is not sufficient. A state could have valid assessments and not have a valid school accountability system. For example, the state's assessments in mathematics may be highly valid and reliable, but the state's use of the assessment data in its accountability system may not be valid or reliable. In addition, most accountability systems incorporate additional indicators beyond assessment outcomes. It is important that a quality assurance plan be developed to validate the use of each measure. As an example, the ASR SCASS recently released a report detailing a quality assurance plan for calculating graduation rates (Taylor, Beaudoin, & Goldschmidt, 2007). The report encourages those working on accountability systems to implement quality checks both at the time of collecting data and when the results are reported. These quality checks begin with training those who are collecting the data, periodic monitoring of the data collection, internal and external audits of the calculations, and statistical verification of the results.

Thus it is important not only to ensure the validity of each indicator in an accountability system, but to ensure the indicators are combined, used, and reported in a meaningful way. These concepts will be discussed in the next two chapters.

References and Suggested Readings for This Chapter

Blank, R. (1993). Developing a System of Education Indicators: Selecting, Implementing, and Reporting Indicators. *Educational Evaluation and Policy Analysis*, 15(1). pp. 65-80.

Council of Chief State School Officers. *State Education Indicators with a Focus on Title I*, (Annual Reports 1997- present). Washington, DC: U.S. Department of Education.

Forte Fast, E. & Hebbler, S. (2004). *A framework for examining validity in state accountability systems*. Washington, DC: Council of Chief State School Officers

Jones, K. (2004). A balanced school accountability model: An alternative to high-stakes testing. *Phi Delta Kappan*, 85(8). pp. 584-590.

Porter, Andrew C. (1991). Creating a System of School Process Indicators. *Educational Evaluation and Policy Analysis*, 13, pp. 13-29.

Raudenbush, R. (2004). *Schooling, statistics, and poverty: Can we measure school improvement?* ETS Angoff Lecture, 2004.

Shavelson, R., McDonanell, L, Oakes, J., Carey, N. (1987). *Indicator Systems for Monitoring Mathematics and Science Education*, Santa Monica, CA: Rand

Taylor, R., Beaudoin, J. & Goldschmidt, P. (2007). *Quality assurance practices associated with producing cohort graduation rates*. Washington, DC: Council of Chief State School Officers.

Checklist for State Accountability Systems: Design Decisions

- *Who is held responsible?*
  - ☐ Develop a system of shared responsibility.
  - ☐ Consider providing information both on students (achievement) and educators (instruction).
- *What accountability model best serves your purpose?*
  - ☐ Deliberately consider status versus improvement versus growth and align the model with the goals.
  - ☐ If the goals imply treating low performing but improving schools differently from low performing schools, ensure accountability system can distinguish among schools that are improving and those that are not.
- *How will data from multiple measures and indicators be combined to categorize schools and make an accountability judgment?*
  - ☐ Make explicit the weighting of elements in the system, including different types of test content and different information sources.
  - ☐ Decide whether your system will use compensatory or conjunctive rules (both at the individual level and in the aggregate) by considering the goals and theory of action.
  - ☐ Consider the use of an index to weight elements appropriately.
  - ☐ Consider the complexity and trade-offs with fairness (complex formulas to create indices can include all relevant factors but be difficult to communicate to the public).
- *What is satisfactory progress?*
  - ☐ Proceed at a pace that is technically defensible and politically reasonable.
  - ☐ Decide whether to hold schools accountable for how they increase student learning or to the absolute level of learning (which includes effects of prior instruction).
  - ☐ Consider holding school systems (e.g., districts) accountable for the cumulative learning of students over their career in the system.
  - ☐ Decide between having the starting point for a school determine the goal or setting the same goal for all schools.
  - ☐ Decide between determining the end point as a fixed point in time, or allowing the annual goals determine the end point for the final goal.
  - ☐ Decide whether each subgroup should meet a separate bar based on their starting point or the same bar.
- *How do we balance reliability and validity concerns in making decisions about schools?*
  - ☐ Strive to make the overall design of the system valid and implement it in a way that leads to valid results. Consider all potential threats to validity and eliminate those as a first step.

- □ Carefully consider the real costs of Type I and Type II errors and attempt to reach a reasonable balance between the two. Do not assume that the costs associated with either type of error are insignificant.
- □ Consider the goal of inclusion of all subgroups when making determinations of a minimum N for subgroups.
- □ Explore the uncertainty involved in measuring school improvement over one year in terms of making reliable judgments for most schools.
- □ Develop rules for determining satisfactory progress of schools and individuals to avoid erroneous judgments attributable to fluctuations of the student population or errors in measurement.

• *How do we move from an old accountability system to a new one?*
- □ Consider previous policies when making new. Decide whether the new system should replace or supplement the old by considering all implications.
- □ Determine how to transition any school categorizations from the old system to the new.
- □ Include a reasonable timeline for the transition.

Once policymakers have decided on a set of indicators, the next question is how to use them to make decisions about teacher, school, or district effectiveness. This issue gets at one of the main sources of discussion of the NCLB regulations—whether school effectiveness is best measured using a status, improvement, or growth model, or some combination of these. Although this issue should have been addressed under goals and used to select indicators, it is discussed here because the model selected relates to how accountability decisions are made. Next, policymakers should decide how to combine the indicators to make decisions about students, teachers, schools, and districts. Will the indicators be combined in a compensatory fashion, where low performance on one measure can be offset by high performance on another or will there be a minimum level of performance set for each measure? Again, the answer to this question will be determined by the goals of the system. In addition, this section addresses decisions regarding school classification, such as how high to set a bar, how often to raise a bar, and how to balance reliability and validity concerns.

Combining Indicators to Make Accountability Decisions

The first design decision involves determining how to combine the different performance indicators to determine if the school has met the goal(s). In a coherent system, many of these decisions will be based on the goals. For instance, weighting the various elements in a system relates directly to the values placed on each element. A different signal is sent if performance indicators are given equal weight as process indicators than if performance indicators are given twice the weight as process indicators.

Another issue is how fine grained the decision measure should be. NCLB breaks down all decisions into "meets annual measurable objectives" or "does not meet annual measurable objectives." However, Hanushek and Raymond (2002) argue that binary pass/fail decisions lead to a set of complication that can be avoided by providing more detailed information about the distribution of scores.

Example: Kentucky

The Commonwealth Accountability Testing System (CATS) goal for every school in the state is Proficiency by 2014 as defined by the Kentucky Board of Education. The goal of Proficiency translates into a school accountability index value of 100. Intermediate targets that will eventually take a school to the goal of 100 are set biennially starting in 2002. As such, there are seven biennia or accountability cycles between 2002 and 2014, as well as recognition points.

Both academic content-based and non-academic measures are used in CATS. These measures include custom, criterion-referenced Kentucky Core Content Tests (KCCT) in reading, mathematics, science, social studies, arts and humanities, practical living/vocational studies and writing, as well as a nationally norm-referenced test in mathematics and reading.

Non-academic measures include attendance rate, retention rate, dropout rate and transition-to-adult-life. Transition-to-adult-life data is collected in the fall of each year via a short survey completed by school personnel. This includes the number of graduates planning to enter the work force, the military, a college or a vocational/ technical school.

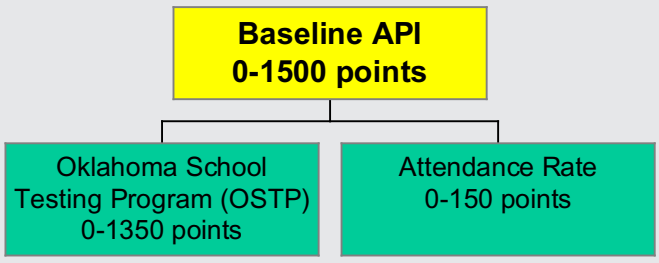
These multiple measures provide a "snapshot" of schools and communicate to schools strengths and weaknesses in order to allocate resources and analyze instructional programs. Each of the measures is combined into a composite to obtain a school's Accountability Index. The CATS goal for every school in the state is Proficiency as defined by the Kentucky Board of Education. The goal of Proficiency translates into a school Accountability Index value of 100. The State goal is for each school to achieve an accountability index of 100 by 2014.

The CATS Accountability Index calculation begins with simple number-correct raw scores and ends with an accountability index that summarizes a school's progress toward the state's goal of Proficiency. Raw scores give rise to scale scores; scale scores have been related to Novice, Apprentice, Proficient and Distinguished (NAPD) performance levels (via standard setting and cut-scores); NAPDs get weighted numerically and combined within each content area; the content areas are weighted and combined to form a school's academic index; and finally, the academic index is combined with the norm-referenced test (NRT) and non-academic factors to generate the accountability index.

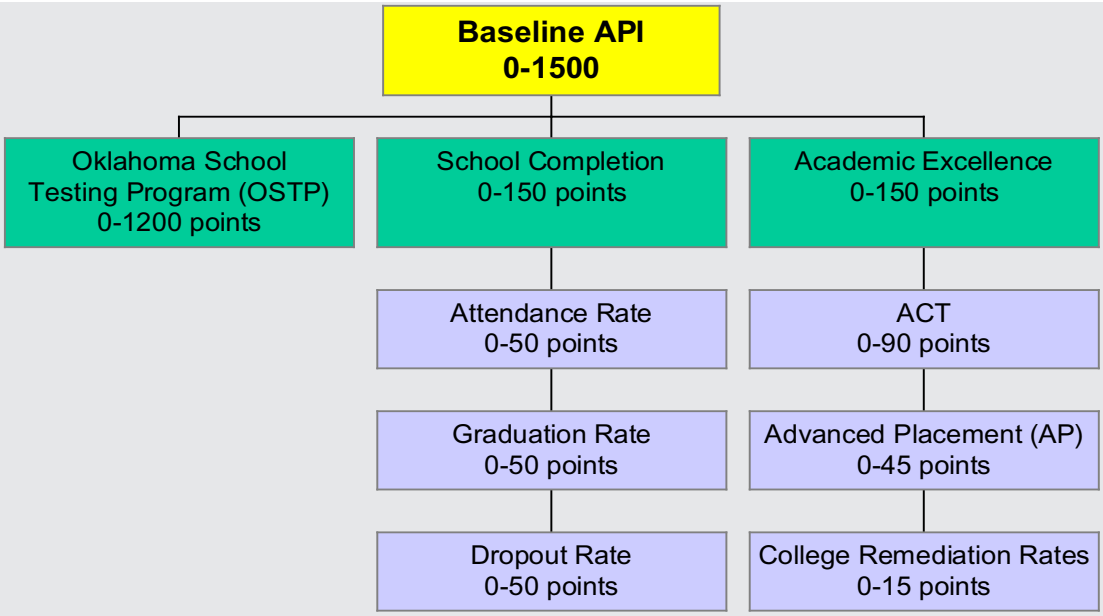
Example: Oklahoma

Oklahoma uses an Academic Performance Index (API). The API is a numeric score that measures school and district performance based on a variety of educational indicators. It allows schools and districts to gauge their progress toward improving student achievement. Seven indicators were mandated to be included in this index, including the Oklahoma State Testing Program (OSTP), ACT, advance placement courses, college remediation, dropout and graduation rates, and attendance. Components of the API are used to meet NCLB reporting requirements. The scale for API score ranges from 0 to 1500; the state average for general education students was 1180 in 2006.

Oklahoma's API: Point Contributions for Elementary and Middle Schools and K-6/K-8 Districts



Oklahoma's API: Point Contributions for High Schools and K-12 Districts



OK state law requires that a minimum 60 percent weight be assigned to the OSTP. After due consideration of the merits of the various combinations, weights were assigned based on site grade span (e.g., elementary) and district type. Table 1 contains the weighting for each component by grade span and district type. For example, in a K-12 district, the OSTP has a weight of 80 percent. The two remaining components, Academic Excellence (B) and school completion (C) are each assigned a weight of 10 percent. So, the API is calculated as: *API = A * .80 + B * .10 + C * .10*

Table 1. OSDE Weights by District (and Site) Type

School Type (Grade Configuration)	Testing Program (OSTP)	School Completion	Academic Excellence	Comment
K-12 District	80%	10%	10%	Base Weight also used for State
K-8 District	90%	10%	n/a	Will have 6 grade levels of OSTP
K-6 District	90%	10%	n/a	Will have 4 grade levels of OSTP
K-6 (elementary)	90%	10%	n/a	Grade 6 may also appear at middle school.
6-8 (middle)	90%	10%	n/a	Grade 6 is sometimes an elementary grade.
9-12 (high school)	80%	10%	10%	Includes EOI examinations (a 7-12 site also has this weight distribution)
K-8 campus are not included as they are generally the same as a K-6 or K-8 district				

Much discussion has been raised under NCLB on whether accountability systems should use status models, improvement models, or growth models. (See Carlson, 2002 for a complete description of the different types of accountability models.) NCLB uses a status model with improvement taken into consideration under safe harbor. The federal government is currently considering adding growth measures for all states and has allowed nine states to use a growth model in a pilot study. But, a state developing its own system will need to determine the appropriateness of the various models under its theory of action. Consider, for example, a school with high performing students, but each year that performance slips a little. The overall performance measure is still above the bar set for the annual goals, but little improvement and slow growth is seen. On the other hand, there is a school that fails to meet the bar each year, but every year more and more students are succeeding and their growth rates are far higher than the district average. Which one of those school is most in need of additional support or should be sanctioned?

Some researchers and policymakers advocate the use of an index to combine the various performance indicators. The index includes the appropriate weights for each component. Examples of performance indices are the API in California and the PPI in Pennsylvania. These two states use their indices very differently, however. In California, the API is used as an additional indicator, meaning a school must make both the AMO and API targets each year to make AYP. The PPI is used as an additional safe harbor for schools in Pennsylvania, meaning that a school can miss its AMO but make its PPI target and still make AYP. Although using indices can help states include all relevant factors in their decision process, it is important to consider the trade-offs with fairness. For example, indices are more difficult to communicate to the public than simple pass/fail decisions based on a single score.

Incorporating Subgroup Information

A major issue in accountability systems these days is how to incorporate information on subgroups. There is strong agreement that they need to be reported separately; otherwise strong majority performance can overshadow the poor performance of a minority. But how do we handle the case of small subgroups?

Under NCLB, we have seen a lot of discussion of "minimum *n*" (i.e., sample size). A large minimum n can increase the reliability of the decisions but because it excludes certain populations from the system, it decreases the validity of the system. Hill and DePascale (2003) have advised policymakers not to set a fixed n for subgroups, because there is no choice of fixed n that will provide for both valid and reliable results, and most choices provide for neither.

Yet the concern still exists: how do we ensure that all students are achieving regardless of their ethnicity, economic status, language background, or disability when the achievement measures are less reliable for small populations? Utah has developed one approach by creating a "super subgroup." This is an aggregate subgroup that includes all students who qualify for a subgroup other than students in the white only subgroup. Individual student proficiencies and progress are added together and divided by the number of students to determine the subgroup proficiency and progress level. The aggregate subgroup calculation allows for maximum accountability because all student subgroups are included; students in a subgroup

with less than 10 students are included in the aggregate calculation. Each student only counts once, regardless of the number of subgroups in which the student qualifies. However, each individual subgroup proficient is reported separately. The use of a super subgroup also eliminates another decision, which is whether to set separate bars for each subgroup depending on their starting point, or whether to require all subgroups to meet the same goal.

Another issue is how to include special populations whose performance may not be reliably measured by traditional assessments, such as students with significant cognitive disabilities or English language learners. All agree that these students should be included in any accountability system, the question is how to measure their performance and growth appropriately. Goertz (2001) described some of the changes that have occurred in measuring these populations in recent years. She focuses on the fact that states are working to increase the inclusion of these students while maintaining the validity of the interpretation of results across all students. She sees states meeting these goals by:

- Testing more students with disabilities
- Researching and using more accommodations
- Developing alternate assessments of achievement for students with disabilities
- Monitoring exclusion rates and incorporating them into accountability systems

Example: Nebraska

Nebraska's system of locally-based classroom assessment makes it possible for all students to be assessed for state reporting within the classroom as part of their regular instruction. If students need accommodations for testing, they are provided exactly the same accommodations they receive for regular instruction. Students with disabilities and those learning the English language are included in state reporting because the classroom assessments are the tools used for the state reporting of all students.

Determining Issues of Timing

There are several issues involving timing. When should the data be collected? When should decisions be made? Should goals be set to increase every year? Every three years? Should there be a timeline for a final goal?

The majority of states give their assessments in the spring but a handful assess in the fall of the next school year. Under one theory, students should be assessed on the content they just learned. However, the concern that many teachers raise is that they lose instructional time at the end of the year between test review and administration time. Moreover, vendors must work quickly to score the tests and report results during the summer. Some states ask for preliminary results back before the school year ends, while others ask for the full reports at the start of the next school year. On the other hand, if states assess students at the beginning of the subsequent year, they are assessing the knowledge retained over the summer, not necessary the knowledge learned in the previous year. Again, scores must be turned around quickly if they are to be of any use to the current year's teachers. Another consideration with timing is also related to item type: open-ended items take longer to score than multiple-choice items. Some states have designed their assessments to include only multiple-choice items to ensure quick turnaround of test results. Others are experimenting with giving different item types at different times or reporting results from different item types at different times.

The second issue related to timing is determining on which data to base school decisions. According to Hill and DePascale (2003), improvement should never be measured over a single year, but should be averaged across years. Several states have chosen to average their data over two years for AYP determinations to mitigate the effects of an anomalous year. Typically, a school may show that they made AYP by showing that either they met their annual measurable objective in the current year or they met in when the previous year's percentages were averaged with the current year's percentages.

Setting Targets

A key feature in any educational accountability system is setting annual targets for students, teachers, and/or schools. Targets are measurable steps towards a system's ultimate goal. But, how do we determine what that target should be; that is, what is good enough? Some researchers argue that targets should be set based on past performance to ensure that the targets are realistic and attainable. The underpinnings of NCLB argue that the desired target of 100% proficiency is attainable in 12 years. Annual targets are determined by calculating the distance between the actual percent proficient in 2002 and the desired percent proficient of 100% of 2014.

Setting targets also incorporates the two topics just discussed – incorporating subgroup information and timing. Policymakers will need to determine if one target should be set for all populations, or if the target should be dependent on starting point. Further, targets may differ for status and growth. Consider, for example, Population A with a current percent proficient of 25% and Population B with a current percent proficient of 50%. Perhaps the status targets are lower for Population A than for Population B over the next 5 years, but the growth are higher. This implies a goal of the two population performing at an equal level at some point in the future, but not immediately.

A related concern in setting targets for satisfactory achievement is the timing for improvement. Under NCLB, 100% proficiency is required by the 2013-2014 school year. Yet, there is no magic to that year. Under other school improvement programs, the end date for 100% proficiency might be based on the starting point. One suggestion by Linn (2006) is to look for the highest rates of improvement seen in any school system across the state and then to set the improvement goal just below that rate. Using existing data about what has been achieved ensures that states set goals that are feasible for their schools.

Finally, once the starting point has been determined, the final goal has been set, and an end date has been chosen, policymakers need to determine how to set intermediate goals. Should the goal for the school increase every year? Every other year? Under NCLB, some states have set annual measurable objectives (AMOs) that increase every year, while others have set their AMOs to increase every three years. Still others set their AMOs to increase ever three years for the first nine years and then every year the next three years, requiring the greatest rate of improvement for schools at the end of the accountability cycle. It seems more likely, however, that the closer a school gets to achieving 100% proficiency, the harder it will become to close that gap.

Examining Technical Issues

An issue with any accountability system is its technical quality. Typically we want to know if the decisions being made are the right ones. That is, if a school is classified as meeting the goal, is it truly succeeding? And are the schools classified as needing improvements really the ones in most need of additional support? We typically talk about the technical issues in terms of the system's reliability and validity. Reliability in its raw form tells us whether schools are ranked in the same order over two occasions. Typically, in NCLB discussions, reliability is discussed in terms of the probability of making correct decisions. So in this case, we look at classification error when examining the reliability.

Validity involves deriving direct inferences from data and has been defined thoroughly in Chapter 2 of Marion et al. (2002). An excerpt of this discussion follows:

> An accountability system can be said to have validity when the evidence is judged to be strong enough to support the inferences that
> - The components of the system are aligned to the purposes, and are working in harmony to help the system accomplish those purposes; and
> - The system is accomplishing what was intended (and did not accomplish what was not intended).

This definition includes three of the critical requirements for the validity of any system:

1. The smooth interaction among the parts of the system, as they are aligned with the system's purposes
2. The outcome or effectiveness of the system in meeting those purposes
3. The breadth and strength of the evidence, and therefore the inference, about the functioning of the system (evidence about both 1 and 2)

The definition also implies that a system can be invalid either because it (a) lacks the evidence to support an inference about its effectiveness, or (b) that the evidence shows that the system is failing to accomplish its goals.

Hill and DePascale (2003) explore issues of reliability and validity in accountability systems and focus on the tradeoffs between the two goals. They conclude that the first step should be to focus on the validity of implementation and use and recommend eliminating all threats to validity. Further, they highlight the importance of using valid assessments, even if that higher validity comes at the cost of lower student-level reliability. They also recommend considering the costs of categorizing a school as meeting the goals when in fact it is a school in need of improvement compared to the costs of categorizing a school as in need of improvement when in fact it is effective. Either of these errors could have significant costs.

References and Suggested Readings for This Chapter

Carlson, D. (2006). *Focusing state educational accountability systems: Four methods of judging school quality and progress*. Dover, NH: National Center for the Improvement of Educational Assessment.

Chester, M. (2005). Making valid and consistent inferences about school effectiveness from multiple measures. *Educational Measurement: Issues and Practice*, 24(4), 40-52.

Goertz, Margaret. (2001). The Federal Role in an Era of Standards-Based Reform. in *The Future of the Federal Role in Elementary and Secondary Education: A Collection of Papers*. Washington, DC: Center for Education Policy.

Hanushek, E. & Raymond, M. (2002). Sorting out accountability systems. In W. Evers & H. Walberg (Eds.) *School accountability* (pp. 75-104). Palo Alto, CA: Stanford University, Hoover Press.

Herman, J. & Haertel, E. (Eds.) 2005. Uses and misuses of data for educational accountability and improvement. 104th yearbook of the National Society for the Study of Education, Part 2. Malden, MA: Blackwell Publishing.

Hill, R. & DePascale, C. (2003). Reliability of No Child Left Behind accountability designs. *Educational Measurement: Issues and Practice*, 22(3), 12-20.

Marion, S., White, C., Carlson, D., Erpenbach, W, Rabinowitz, S., & Sheinker, J. (2002). *Making valid and reliable decisions in determining adequate yearly progress*. Washington, DC: Council of Chief State School Officers.

Peterson, P. E., and West, M. R. (2006). Is your school effective? *Education Next* (Vol. 2006 No. 4).

Chapter 5: Consequences

Checklist for State Accountability Systems: Consequences

- What consequences (positive and negative) can your accountability system support?
  - □ Determine the appropriate consequences for your goals (remember that providing appropriate support in response to an identified need is a consequence).
  - □ Begin with broad, diffuse stakes and move to specific consequences for individuals and institutions as the system aligns.
- What rewards & sanctions are effective in obtaining the goals of your accountability system?
  - □ Base individual and collective stakes on defensible, empirically-based theories about what is possible to accomplish on measured performance within a given period of time.
  - □ Coordinate stakes for accountability systems to support system goals.
- Who will the consequences affect?
  - □ Consider student incentives (is it fair to have consequences that affect teachers if students are not motivated to do their best?).
  - □ Consider separate consequences for students, educators, schools, and districts.
- When and how will the consequences be applied?
  - □ Clearly state your decision rules and the order and timing of consequences.
  - □ Make appeal procedures available to ensure valid and reliable decisions.
- How will you ensure the consequences are applied appropriately and effectively?
  - □ Make explicit the stakes for results and their phase-in schedule at the outset of the implementation of the system.
  - □ Create stability and credibility (educators must believe rewards and sanctions will happen).
- How will you determine if the consequences are effective?
  - □ Monitor the effectiveness of any rewards (e.g., incentives for teachers) or sanctions (e.g., restructuring schools) on performance.

Along with developing goals, policymakers must implement consequences surrounding those goals. Consequences could be positive or negative. That is, in most accountability models schools that meet the goals are rewarded and schools that fail to meet the goals are sanctioned or receive some type of intervention or support. Rewards could be tangible such as monetary rewards being given to all teachers who raise their students' test scores by a certain amount. They can also take the form of public recognition. The key elements on consequences focus on determining appropriate consequences, targeting them to the appropriate people, applying the effectively, and monitoring the results of these consequences on student achievement and other less tangible outcomes, such as teacher morale and student behavior characteristics.

There is little research that provides us with an answer to the question of which rewards or sanctions are most effective. Hanushek and Raymond (2002) argue that different people react to consequences differently. Some are more motivated by rewards; others are more motivated to avoid sanctions. So, accountability systems may need to include both rewards and sanctions.

In an Education Week article, Olson (2006) reports that few schools in the "corrective action" category are actually being sanctioned by their district. The article cites concerns that some NCLB sanctions may conflict with state law or union employee rights. It suggests another motivator may be that district policymakers want to avoid animosity and thus choose to implement supports rather than sanctions to schools that are not meeting the goals.

Another finding is that the effectiveness of sanctions and rewards is dependent on the stability and credibility of these policies. That is, educators must believe that rewards and sanctions will actually happen (Fuhrman, 1999). She goes on to say that her model suggests that teachers will be motivated to reach performance goals to the extent they

1. perceive a high probability that teacher effort will result in students meeting goals (the clearer and more specific the goals, the more likely the teachers will think their efforts will lead to reaching them),
2. perceive a high probability that goal attainment will lead to consequences or outcomes such as a reward, and
3. place value on these outcomes.

Rewards

Some states use monetary rewards for high performance. For example, Oklahoma recently passed legislation to provide two types of academic achievement awards based on schools scores on the state index:

- Schools with the highest API score or greatest gain in API score by school size (5 categories) for all students will receive monetary rewards
- Schools with a perfect 1500 API score for regular education students will receive monetary rewards

California also used a rewards system in 2000 and 2001. This system provided $750,000,000 to schools and staff, allocated based on the state accountability system (API). Individual teachers could receive up to $25,000. However, California had to abandon this program after 2001 because of fiscal concerns.

Several other states also provide monetary rewards to schools who meet their goal. In fact a 2005 review by the Education Commission of the States found that at least 30 states offer financial incentives to raise student scores, sometimes in the form of student loan forgiveness, housing benefits, or scholarships, in addition to bonuses and raises.

There is some evidence that monetary rewards are associated with improvement in student achievement data. For example, one study in Kentucky showed that student test scores increased between 1992 and 1998 when a school-based performance award program was

implemented. The greatest gains occurred during the first two years of the program. Kelley (2002) describes conditions that must be in place for these financial incentive programs to be successful. The programs must provide clear goals for the teachers and must be perceived as fair in order for them to show results. In addition, the system should be designed so that the incentives encourage desirable behaviors—that is increasing student learning is desirable, while encouraging cheating is not. Systems that promote whole-school rewards encourage teachers to work together and problem solve collaboratively.

However, there are several concerns with monetary rewards. One of which is that if they reward is promised it must be given, and in a timely manner. There are several examples of states promising monetary incentives to teachers who raise test scores but not delivering any funds when those goals were actually reached. Secondly, there is a concern that monetary rewards are less motivating than appropriate support or more intrinsic rewards. Teachers rarely go into the profession for the money; most naturally want to see student perform to their potential. Therefore a reward more in line with their beliefs would be recognition of the students and their teachers.

As one model, rewards could take the form of mentoring, professional development, and updated technology or additional supplies for the classrooms. Another model involves public recognition of teachers, schools, and districts who are raising student achievement and retention. As one example, Nebraska publishes an honor on its web site for those who have both high quality local assessments as determined through the validation/verification process as well as high student achievement.

Other states approach a reward as an avoidance of a sanction. In Missouri, for example, school districts are audited on a five-year cycle to keep their accreditation. Districts that demonstrate consistently high performance among all student groups may have that audit waived. Schools can also be designated as A+ or Exemplary and can waive audits as well.

Intrinsic Rewards

Nebraska asserts that the real "rewards" have come through the professional development of teachers and their knowledge of their own growth. For instance, one teacher commented about her local accountability system, "I guess I think it's made me a better teacher. And I think after teaching for a number of years, we tend to get in a rut and sometimes think we've got it all down; we know what we're doing and I think we need to be reminded that there's always room for us to learn." Another teacher focused on her understanding of incorporating the test objectives into the curriculum: "Our staff is light years ahead of where they used to be. I think they think more about planning backwards. In other words, they say, what are we assessing and then how do we get our students to that? I think they're more concerned with what they're assessing and how students are progressing."

Sanctions

NCLB mandates specific sanctions for schools who do not meet AYP for two or more consecutive years. Recent research has provided some insight into the efficacy of these sanctions. A Title I study commissioned by the U.S. Department of Education and conducted by

RAND provides evidence that providing tutoring is more effective in raising student achievement than allowing students to transfer from a school in improvement status to one that is meeting AYP (Zimmer, et al., 2007).

One area where research shows agreement is that policymakers should monitor the responses of schools and educators to these sanctions and rewards (cf., Elmore, 2004 or Hamilton & Stecher, 2004). Because there is insufficient information on which sanctions and rewards are most effective, it is important to monitor the effects of consequences on student performance. In addition, possible negative effects on other aspects of schools should be monitored, such as lower teacher morale, increased teacher turnover, increased administrator turnover, narrowing of the curriculum, or the elimination of electives such as art and music.

Some states have identified other negative consequences such as schools "blaming" specific student groups for the school's failing to meet its annual objectives. For example, a school principal may be less welcoming to new special education students if the school is failing to make its annual objectives only in the special education population. Another concern in some states is that the focus on Title I schools has led to school improvement supports being implemented only in Title I schools—or only Title I schools that fail to make AYP—even though all schools could benefit from the support.

Example: Kentucky

Kentucky's Commonwealth Accountability Testing System (CATS) establishes a biennial system of rewards for school improvement and sanctions for schools that do not perform as expected. These sanctions exist in the form of assistance. According to regulation, all schools falling into the Assistance classification will be rank-ordered from highest to lowest according to the school's combined Accountability Index. This set of schools will then be divided into thirds. The top third will be designated Level 1 schools, the middle third Level 2, and the bottom third Level 3. The following bullets briefly summarize the audit/review process for these schools:

Schools in Assistance are ranked and grouped into three levels:
Level 1: scholastic self-review
Level 2: Scholastic review
Level 3: Scholastic audit

Level 1—The school shall adhere to the requirements for a "Level 1" school as defined in 703 KAR 5:120 Section 2. Level 1 schools must conduct a scholastic review and self-study facilitated by the district's professional development coordinator with assistance provided by Kentucky Department of Education staff. Assistance Level 1 schools may be eligible to receive Commonwealth school improvement funds.

Level 2—The school shall adhere to the requirements for a "Level 2" school as defined in 703 KAR 5:120 Section 3. Schools are required to receive a scholastic review by a team set up by KDE. The team must include local district members. Level 2 schools shall receive a scholastic review facilitated by a designee of the Commissioner of Education with assistance from the district's central office staff. Assistance Level 2 schools may be eligible to receive Commonwealth school improvement funds.

Level 3—The school shall adhere to the requirements for a "Level 3" school as defined in 703 KAR 5:120 Sections 4, 5, 6, 7, 8 and 9. Schools will be scheduled for scholastic audits by an external team coordinated by KDE. Level 3 schools shall receive education assistance from a highly skilled educator under KRS 158.782 and a scholastic audit. Assistance Level 3 schools may be eligible to receive Commonwealth school improvement funds.

References and Suggested Readings for This Chapter

Elmore, R. (2004). Conclusion: The Problem of Stakes in Performance-Based Accountability Systems. In Fuhrman, S. & Elmore, R., (2004). *Redesigning Accountability Systems for Education*. New York, NY: Teachers College Press.

Fuhrman, S. (1999, January). The new accountability. *CPRE Policy Briefs*, RB-27. Philadelphia, PA: The Consortium for Policy Research in Education.

Hamilton, L. & Stecher, B. (2004). Responding Effectively to Test-Based Accountability. *Phi Delta Kappan*, 85(8). pp. 578-583.

Hanushek, E. & Raymond, M. (2002). Sorting out accountability systems. In W. Evers & H. Walberg (Eds.) *School accountability* (pp. 75-104). Palo Alto, CA: Stanford University, Hoover Press.

Kelley, C. (2002) Financial incentives in state accountability systems: Performance pay for teachers. Madison, WI: Consortium for Policy Research in Education.

Olson, L. (2006, December 6). U.S. urged to rethink NCLB "tools": Districts seen as using light touch for schools required to restructure. *Education Week*, pp. 1, 19.

Rustique-Forrester, E. (2005). Accountability and the pressures to exclude: A cautionary tale from England. *Education Policy Analysis Archives*, 13(26). Retrieved February 13, 2007 from http://epaa.asu.edu/epaa/v13n26/.

Zimmer, R., Gill, B., Razquin, P., Booker, K., & Lockwood, J.R. (2007). *State and local implementation of the No Child Left Behind Act: Volume I—Title I school choice, supplemental education services, and student achievement*. Washington, DC: U. S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.

Chapter 6: Communication

Checklist for State Accountability Systems: Communication

- *How will accountability results be reported?*
  - □ Ensure the reports to districts and schools promote appropriate interpretations and use of results.
  - □ Provide data to schools in a way that they can learn to use it.
  - □ Monitor the latest research in effective reporting and continually review the reporting system.
  - □ Ensure reports clearly communicate results and include all relevant data.
  - □ Include multiple indicators of performance broken out by subgroups in the reports.
  - □ Make error estimates of all measures available.
  - □ Ensure all reports meet FERPA regulations (e.g., do not report individually identifiable data)
  - □ If there are consequences associated with the results, the results must be provided to local school systems with time for appeal before results are released to the general public.
- *How will parents and the general public be informed as to the goals and limitations of the system?*
  - □ Make accountability expectations public and understandable for all participants in the system.
  - □ Prepare information guides specific to the audience with various levels of detail as appropriate.
  - □ Make system results broadly available to the press, with sufficient time for reasonable analysis and with clear explanations of legitimate and potential illegitimate interpretations of results.

The next important concept is communication, which includes both communication about the accountability system, its goals and consequences at the inception, and communication of results, such as score reporting. The key elements focus on providing data to stakeholders and the general public in a manner that is both understandable and useful. These data include both information on the accountability system itself as well as the results.

When we think of communication within an accountability system, we typically think first of reporting the results. However, beyond reporting the results, an important component of an effective accountability system involves communicating the goals, the mechanisms through which we expect to implement change, how this change will be monitored and evaluated, and the consequences for meeting or not meeting the goals. For instance, Hamilton and Stecher (2004) indicate that to improve instructional outcomes, the standards must be fully communicated to educators and parents. In order to effect change, educators and parents must understand what is valued. That is, they need to know what the goals are and how they will be measured.

We also need to consider communication within the system on an ongoing basis. O'Day (2002) analyzed accountability systems more broadly and found a key component of success was the ability of a system to "communicate across units." Applying this idea to school accountability, we can see the importance of communicating from the state to the district to the school and back up again. We can also see the importance of communicating within the school to ensure a coherent system of instructional improvement both across departments and engaging the home environment through involving the parents.

In her article discussing the lack of sanctions imposed on schools, Olson (2006) notes that one issue that appears to be reducing participation in the school choice option under NCLB is poor communication with the parents. The letters that go out to parents explaining that their child's school has been placed in a corrective action category are often quite long and full of legal jargon. If the parent does not understand their options, they are unlikely to choose any of them.

When reporting results, it is worth noting that there is a type of accountability system based on performance reporting. Performance reporting assumes that collecting and publicly reporting information on student academic performance, combined with rewards or sanctions, will result in improvements in teaching and learning. However, the early literature showed no direct link between performance reporting and school improvement (Kirst, 1990). One issue that is still apparent today is the need to develop strong content standards, performance standards, and reports that provide information in a way that educators are able to use it. Timing is also key. The scores should be reported in a timely manner that allows instructors to relate the results to their instructional practices. Fuhrman (1999) notes that public reporting is a shift in educational accountability that came about in the 1990s. There has been an increase in the study of how best to communicate test results over the last 15 years (cf., Forte Fast & ASR SCASS, 2003).

Baker and Linn (2002) indicate that two of the four conditions required of an effective accountability system involve communication of the results: (1) the results reported are accurate and (2) the results are validly interpreted. Carlson (2006) and Erpenbach (2002) also comment on the importance of publicly reporting the results with the emphasis on public. The Baker et al. (2002) policy brief devotes an entire category to public reporting. Their two primary standards are:

1. System results should be made broadly available to the press, with sufficient time for reasonable analysis and with clear explanations of legitimate and potential illegitimate interpretations of results.
2. Reports to districts and schools should promote appropriate interpretations and use of results by including multiple indicators of performance, error estimates and performance by subgroups.

These standards reiterate that those responsible for the accountability system also "have a responsibility to help ensure proper interpretation of the results and to minimize inappropriate interpretations to the extent possible" (Baker et al., 2002, p. 5). The authors focus on the press, because the press plays an important role in the interpretation of the results. They go on to emphasize that policymakers should devote some effort in assisting the press in under-

standing the technical characteristics of the accountability system and its results including its strengths and limitations. Hamilton and Stecher (2004) assert that the mechanism for improving instruction through assessment requires that educators be trained in interpreting and using the results effectively. They also focus on the role of score reports in providing to parents information that they need to make better decisions for their children.

Furthermore, Baker et al. (2002) recommend analyzing the overall picture by interpreting and reporting the consistencies and inconsistencies provided by the multiple indicators of performance. They further recommend explaining the error involved in any measure and helping the reader understand the degree of uncertainty in results to reduce the likelihood of misinterpretation and increase the likelihood of appropriate use of results.

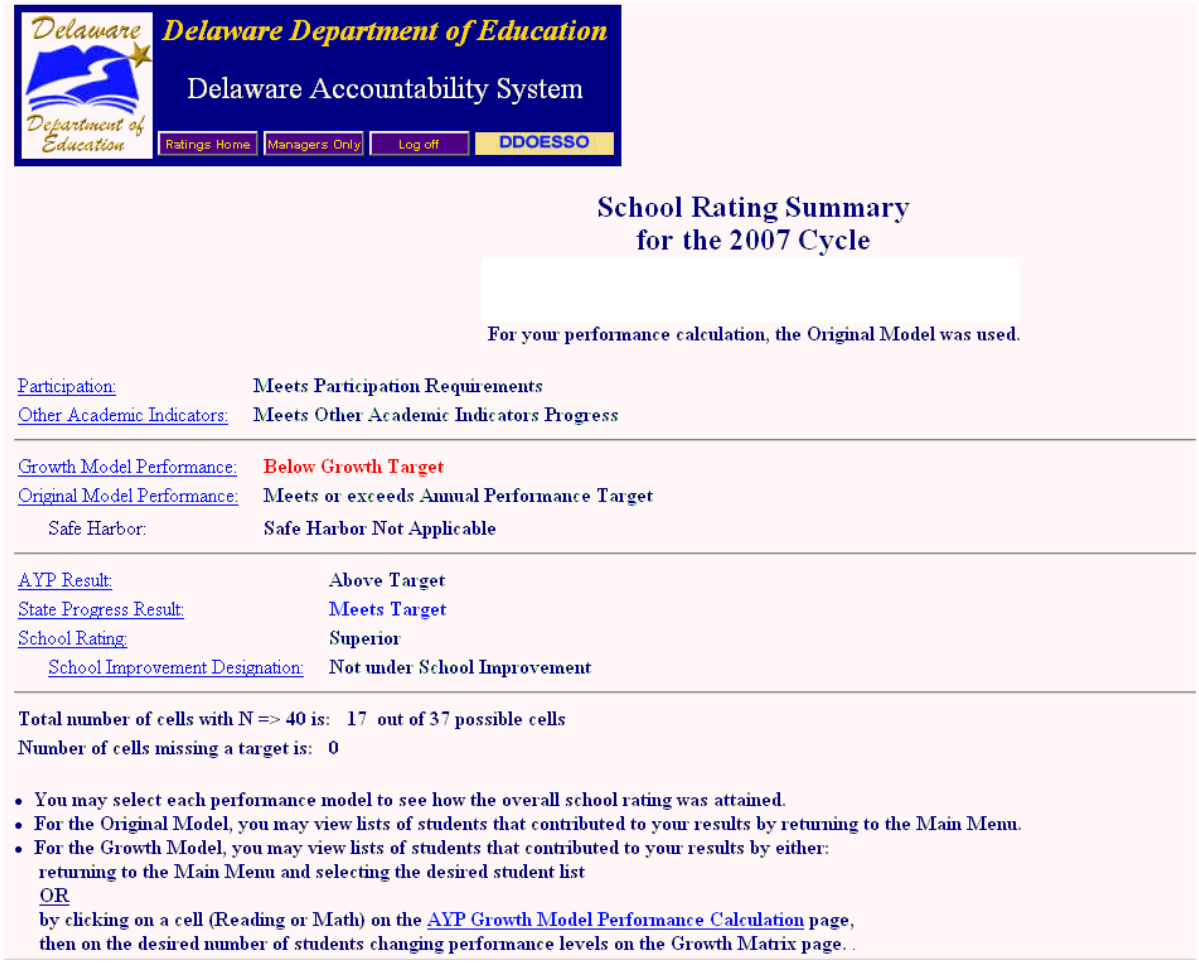
The focus on reporting subgroup information is also supported by the Standards for Educational and Psychological Testing (APA, AERA, & NCME, 1999):

> *Standard 7.8: When scores are disaggregated and publicly reported for groups identified by characteristics such as gender, ethnicity, age, language proficiency, or disability, cautionary documents should be included whenever credible research reports that test scores may not have comparable meaning across those different groups.*

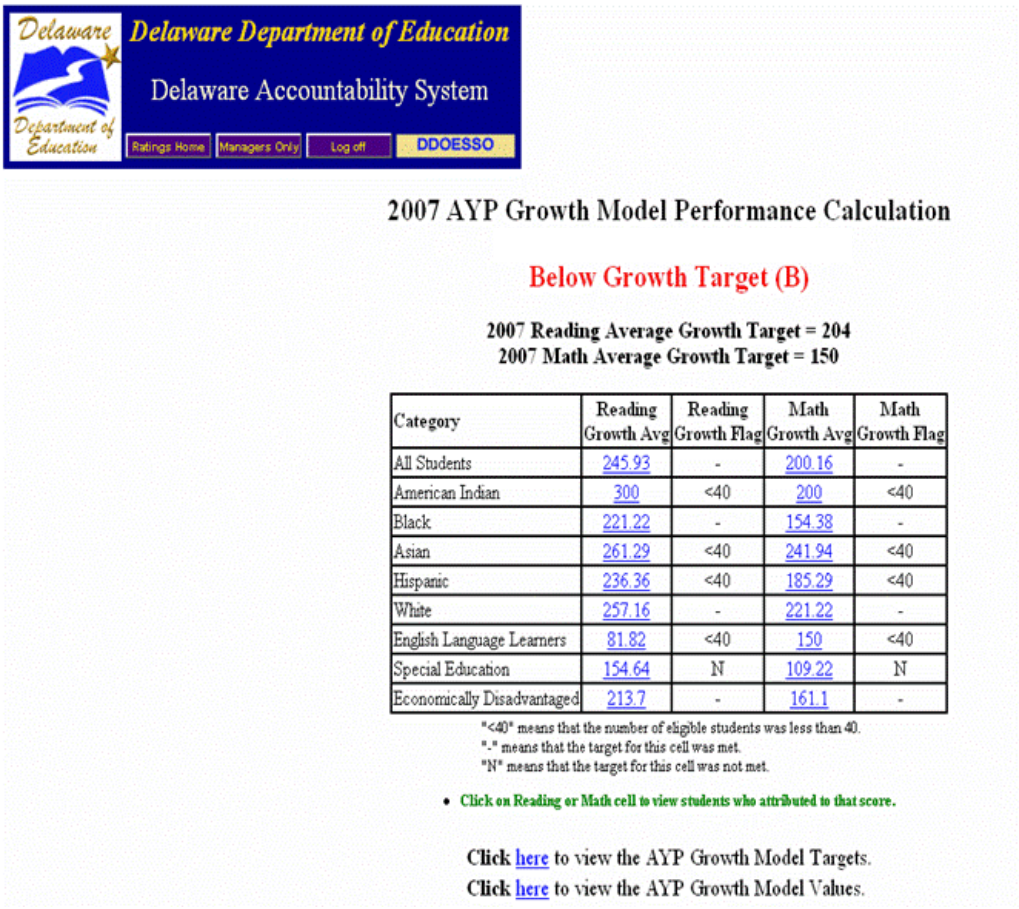
Baker et al. (2002) comment that analyzing performance by subgroups helps ensure that disparities in subgroups are not hidden by the overall results. That is, if one subgroup comprises a small percentage of the population, their poor performance may be concealed by strong performance of the majority subgroup.

Many states have developed online reporting systems over the past several years to provide more timely and detailed information to schools, teachers, and the community. For example, Delaware created an online reporting system in 2003 that is enhanced annually. It includes multiple years of detailed information, including scores and the numbers behind the scores. Educators can access state scores, district scores, school scores, student scores, and subgroup scores. Information on students over the years allows educators to track individuals' growth.

See the box below for an example of the information provided for a single school in Delaware. Educators within that school can drill down to see further detail for each level of reporting.



As a further example, we can drill down into the Growth Model Performance to determine why this school was labeled as "Below Growth Target." The next screen, shown below, shows the target for this school in reading and math and then shows the actual scores for each subgroup for each subject. From this chart, we see that the special education students in this school failed to meet the growth target in both reading and math. English language learners also failed to meet the target in reading, but they also did not meet the minimum number of eligible students, so this score did not count against this school.



Finally, educators can drill down to see the formulas to determine how the composite scores were calculated. They are given the formula:

Composite Score = 25 [(25% X Reading Score) + (25% X Math Score) + (25% X Science Score) + (25% X Social Studies Score)]

And then shown how the formula applies to their data.

Your Content Scores were:

The Reading Score: 2.91 = (5 x (7/432)) +
(4 x (39/432)) +
(3 x (307/432)) +
(2 x (68/432)) +
(1 x (11/432))

The Math Score: 3.07 = (5 x (80/432)) +
(4 x (68/432)) +
(3 x (154/432)) +
(2 x (62/432)) +
(1 x (68/432))

The Science Score: 2.94 = (5 x (28/347)) +
(4 x (56/347)) +
(3 x (160/347)) +
(2 x (74/347)) +
(1 x (29/347))

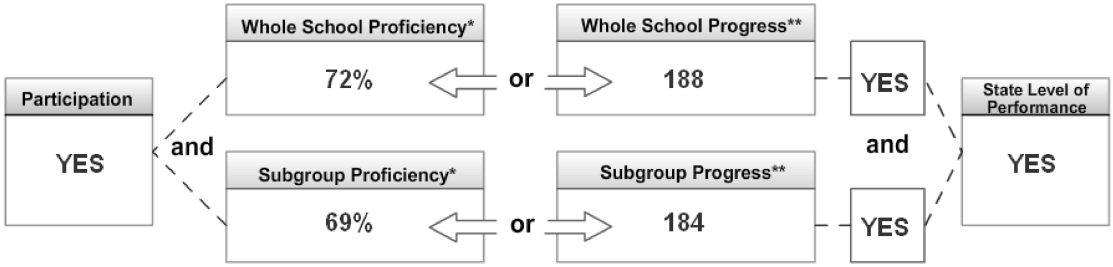
The Social Studies Score: 2.80 = (5 x (30/346)) +
(4 x (43/346)) +
(3 x (137/346)) +
(2 x (99/346)) +
(1 x (37/346))

CompositeScore: 73.27 = 25x[(2.91x.25) + (3.07 x .25) + (2.94x.25) + (2.80 x .25)]

Actual calculations are performed to more decimal places than shown. A cumulative rounding difference is possible using the values shown.

Many states have developed similar online systems. Some of these systems include features that allow states to create their own reports to print. Others include tools to allow educators to focus on specific statistics and develop improvement plans around them. As another example, Utah reports data on their UPASS accountability system in a manner that shows how decisions were made and uses color to indicate areas in which each school succeeded and failed.

U-PASS Accountability System
2006 School Report
North Star School
Salt Lake District



Proficiency		Progress	
LANGUAGE ARTS Percent Proficient	72% Click here for more info.	LANGUAGE ARTS Progress Score	190
MATH Percent Proficient	73% Click here for more info.	MATH Progress Score	192
SCIENCE Percent Proficient	61% Click here for more info.	SCIENCE Progress Score	182
ATTENDANCE Percent Proficient	86%	ATTENDANCE Progress Score	175

In order to achieve the State Level of Performance a school must have 95% participation, and either proficiency or progress in the Whole School and the Subgroup.

*Overall proficiency:
Acceptable range is 80 % and higher.

** Overall Progress:
Acceptable range is 190 and higher.

Participation:
Acceptable range is 95% and higher.

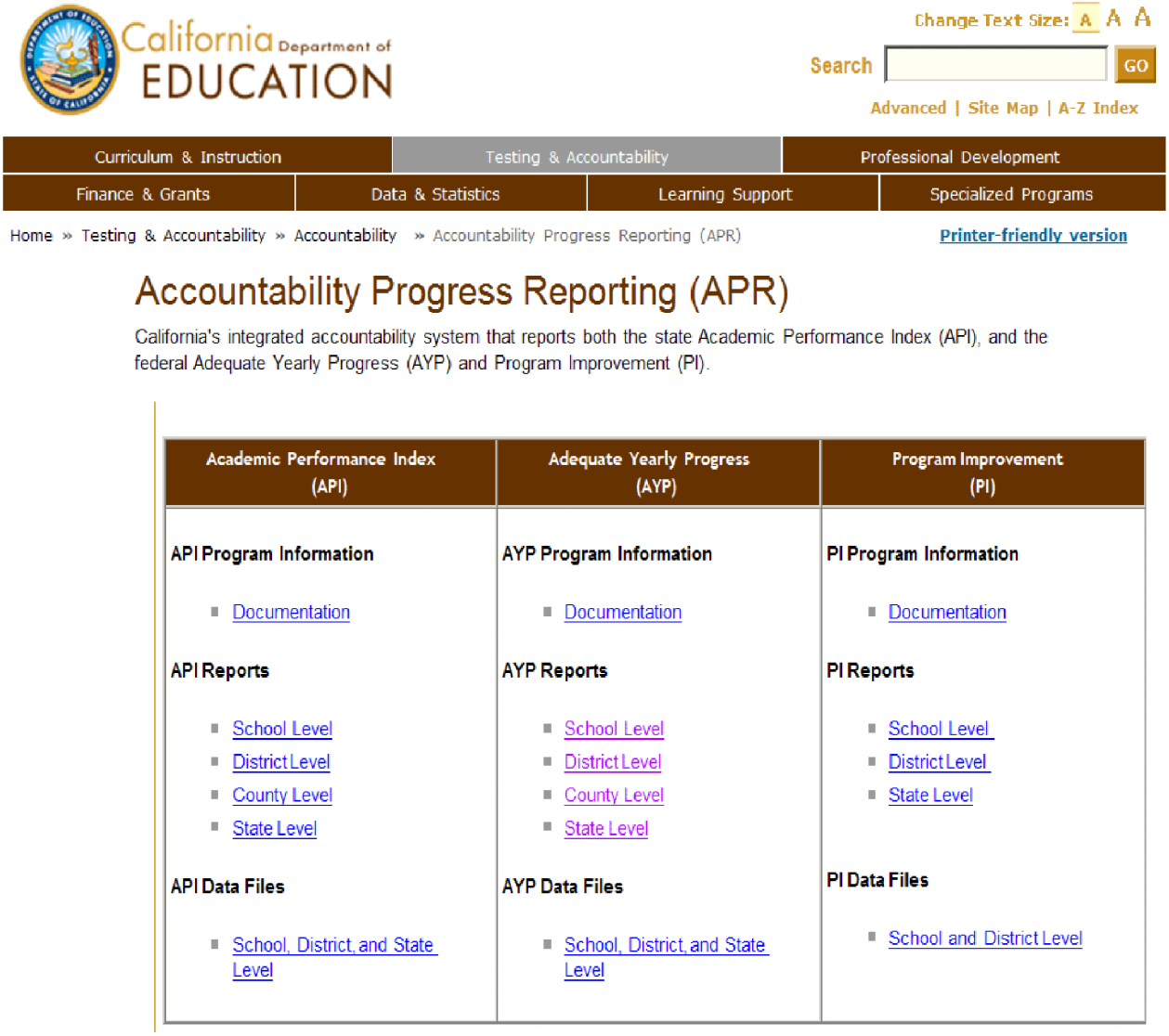
Click here for additional school information

Individual Subgroups	Proficiency	Progress
African American	yes	n<10
American Indian	n<10	n<10
Asian	yes	yes
Caucasian	yes	yes
Hispanic	no	yes
Pacific Islander	no	yes
Economically Disadvantaged	no	yes
Limited English Proficient	no	yes
Students With Disabilities	no	no

Communicating Results from Dual Accountability Systems

Linn (2005b) has completed several analyses of states with dual accountability systems (e.g., AYP and a state system). In such states, communication becomes even more crucial as educators and parents often receive mixed messages from the two systems. For instance, how should a parent in Colorado evaluate their child's school when they are told that it is a high performing school that failed to make AYP? The two systems may have different sanctions attached to failing to meet the goals, and these must be clearly communicated to those affected by such sanctions.

A different example is provided by California, which maintains a dual accountability system. While they also have an extensive online reporting system, they provide detailed information on the two systems and report results separately for each. For example, the following is their home page for accountability reporting:



California Department of EDUCATION

Change Text Size: A A A

Search GO

Advanced | Site Map | A-Z Index

Curriculum & Instruction | Testing & Accountability | Professional Development

Finance & Grants | Data & Statistics | Learning Support | Specialized Programs

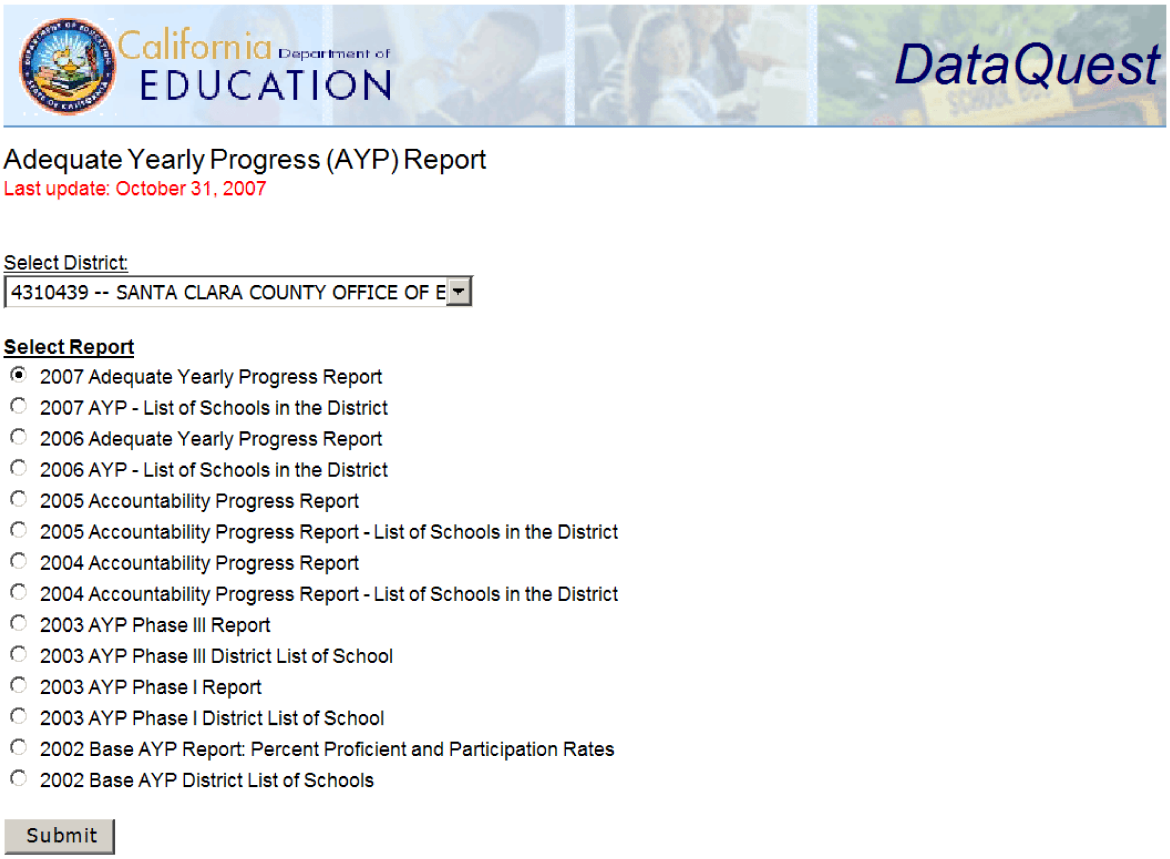
Home » Testing & Accountability » Accountability » Accountability Progress Reporting (APR) Printer-friendly version

Accountability Progress Reporting (APR)

California's integrated accountability system that reports both the state Academic Performance Index (API), and the federal Adequate Yearly Progress (AYP) and Program Improvement (PI).

Academic Performance Index (API)	Adequate Yearly Progress (AYP)	Program Improvement (PI)
API Program Information ▪ Documentation	**AYP Program Information** ▪ Documentation	**PI Program Information** ▪ Documentation
API Reports ▪ School Level ▪ District Level ▪ County Level ▪ State Level	**AYP Reports** ▪ School Level ▪ District Level ▪ County Level ▪ State Level	**PI Reports** ▪ School Level ▪ District Level ▪ State Level
API Data Files ▪ School, District, and State Level	**AYP Data Files** ▪ School, District, and State Level	**PI Data Files** ▪ School and District Level

Each system has its own documentation section that includes an overview of the system, related legislation and references, parent guides in both English and Spanish, technical specifications, calculation spreadsheets, data files, and reports. Reports are provided at the state, county, district, and school level. Clicking on the district level link under API reports and entering the name of a district produces the following screen:



Clicking on that first radio button for this district, then selecting the 2007 Growth heading produces the following screen:

2006-07 Accountability Progress Reporting (APR)

Local Educational Agency (LEA) Report - Growth API

2007 Growth Academic Performanc Index (API) Report

California Department of Education
Policy and Evaluation Division
11/5/2007

LEA: Santa Clara Unified
LEA Type: Unified
County: Santa Clara
CD Code: 43-69674

2007 Growth API Links:

- LEA Chart
- LEA Demographic Characteristics
- LEA Content Area Weights
- LEA List of Schools
- County List of Schools

(An LEA is a school district or county office of education.)

2006-07 APR		2006-07 State API				2007 Federal AYP and PI		
Summary	Glossary	2006 Base	Guide	2007 Growth	Guide	AYP	PI	Guide

State Accountability: Academic Performance Index (API)

Number of Students Included in the 2007 Growth API	API		
	2007 Growth	2006 Base	2006-07 Growth
8,971	755	747	8

Growth API target information is not applicable to LEAs or to schools in the Alternative Schools Accountability Model (ASAM).

	Number of Students Included in 2007 API	Numberically Significant in Both Years	Subgroup API 2007 Growth	2006 Base	2006-07 Growth
Subgroups					
African American (not of Hispanic origin)	509	Yes	671	666	5
American Indian or Alaska Native	55	No			
Asian	1,856	Yes	862	850	12
Filipino	719	Yes	782	771	11
Hispanic or Latino	2,726	Yes	657	652	5
Pacific Islander	108	Yes	700	702	-2
White (not of Hispanic origin)	2,938	Yes	790	784	6
Socioeconomically Disadvantaged	3,439	Yes	664	661	3
English Learners	3,097	Yes	717	700	17
Students with Disabilities	1,315	Yes	568	522	46

Click on the column header to view notes.

In order to meet federal requirements of No Child Left Behind, a 2007 Growth API is posted even if a school or LEA had no 2006 Base API. However, the presentation of actual growth would not be appropriate and, therefore, is omitted.

Direct-funded charter schools are not included in the LEA Report.

"N/A" means a number is not applicable or not available due to missing data.
"*" means this API is calculated for a small LEA, defined as having between 11 and 99 valid Standardized Testing and Reporting (STAR) Program test scores included in the API. APIs based on small numbers of students are less reliable and therefore should be carefully interpreted.

The federal Adequate Yearly Progress (AYP) requirement for the API is: a 2007 Growth API of 590 OR a one-point increase from the 2006 Base API to the 2007 Growth API for a school or LEA.

Using the bars in the upper part of the screen, a user can toggle between the full accountability results, the API (state) results, and the AYP (federal) results. The user can also click on one of the bars in the upper right part of the screen to get a list of schools in this particular district and determine the API and AYP status of each.

Clicking on the summary button on the left-hand side provides the user with the overall status for this particular district. The following screen shot shows that this district's API score grew, it made AYP, and it is not in program improvement.

2006-07 Accountability Progress Reporting (APR)

Local Educational Agency (LEA) Summary
2006-07 APR

California Department of Education
Policy and Evaluation Division
11/5/2007

LEA: Santa Clara Unified
LEA Type: Unified
County: Santa Clara
CD Code: 43-69674

2006-07 APR Links:

- Base API LEA List of Schools
- Base API County List of Schools
- Growth API LEA List of Schools
- Growth API County List of Schools
- AYP LEA List of Schools
- AYP County List of Schools

(An LEA is a school district or county office of education.)

2006-07 APR		2006-07 State API				2007 Federal AYP and PI		
Summary	Glossary	2006 Base	Guide	2007 Growth	Guide	AYP	PI	Guide

State Accountability: Academic Performance Index (API)

2006 Base API	2007 Growth API	Growth in the API from 2006 to 2007
747	755	8

Growth API target information is not applicable to LEAs, to schools in the Alternative Schools Accountability Model (ASAM), or to schools that do not have a valid 2006 Base API.

Federal Accountability: Adequate Yearly Progress (AYP)

Made AYP: No

	English-Language Arts	Mathematics
Met AYP Criteria:		
Participation Rate	No	Yes
Percent Proficient	Yes	Yes
API - Additional Indicator for AYP	Yes	
Graduation Rate	Yes	

Program Improvement (PI)

PI Status: Not In PI

References and Suggested Readings for This Chapter

American Psychological Association, American Educational Research Association, & the National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Baker, E. & Linn, R. (2002). *Validity issues for accountability systems* [Technical Report 585]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing

Baker, E., Linn, R., Herman, J., & Koretz, D. (2002). *Standards for educational accountability systems* [Policy Brief 5]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing

Carlson, D. (2006). *Focusing state educational accountability systems: Four methods of judging school quality and progress*. Dover, NH: National Center for the Improvement of Educational Assessment.

Erpenbach, W. (2002). *Incorporating multiple measures of student performance into state accountability systems—A compendium of resources*. Washington, DC: Council of Chief State School Officers.

Forte Fast, E. & ASR SCASS. (2003). *A guide to effective accountability reporting*. Washington, DC: Council of Chief State School Officers.

Fuhrman, S. (1999, January). The new accountability. *CPRE Policy Briefs*, RB-27. Philadelphia, PA: The Consortium for Policy Research in Education.

Hamilton, L. & Stecher, B. (2004). Responding Effectively to Test-Based Accountability. *Phi Delta Kappan*, 85(8). pp. 578-583.

Kirst, Michael W. (1990). *Accountability: Implications for State and Local Policymakers*. Washington, DC: Information Services, Office of Educational Research and Improvement, U.S. Department of Education.

Linn, R. (2005b). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *Educational Policy Analysis Archives*, 13(33). Retrieved January 15, 2007 from http://epaa.asu.edu/epaa/v13n33.

O'Day, J. (2002). Complexity, accountability, and school improvement. *Harvard Educational Review*, 72(3), pp. 293-329.

Chapter 7: Supports

Checklist for State Accountability Systems: Support

- How will the state provide the schools with the support needed to improve the results required in the accountability model?
  - □ Acknowledge the reciprocity of accountability and capacity.
  - □ Develop an infrastructure to provide expertise and support at local levels.
  - □ Create incentives to stabilize the population of teachers and administrators in low-performing schools.
  - □ Encourage school districts to develop systemwide improvement strategies.
- What are the necessary tradeoffs between quality and cost?
  - □ Determine the actual costs incurred by the system.
  - □ Allocate resources where they are needed most.
  - □ Consider state capacity for remedy.
- What support do administrators need to implement the accountability system?
  - □ Develop plan for supports.
  - □ Communicate clear, high expectations to students.
  - □ Provide resources for professional development aligned to district's mission.
  - □ Consider diagnostics for students who are not succeeding.
- What support (e.g., professional development) do teachers need to implement the system?
  - □ Align professional development to district's mission.
  - □ Provide teachers with sufficient time for planning.
  - □ Provide students with the opportunity to learn.
  - □ Implement appropriate instructional interventions.
- To what degree does the system support high-quality instruction and student access to education?
  - □ Collaborate with appropriate stakeholders to create conditions for a school or school district's success.
  - □ Consider involving schools and school districts in collecting and analyzing information about the effectiveness of educational programs.
  - □ Document and disseminate success stories where high-quality instruction and student access to education were maximized.

The sixth core concept focuses on supports for schools and districts as they try to attain the goals of the accountability system. The key elements focus on the roles of the state, district, and school agents in developing a plan for school improvement, communicating this plan, and providing the necessary resources to ensure that each school can meet the overarching goals.

Newmann, King, and Rigdon (1997) state that "External accountability alone offers no assurance that a school faculty will have adequate technical knowledge and skill, sufficient

authority to deploy resources wisely, or shared commitment to a clear purpose for student learning." In fact, they argue that to improve school quality it needs both the organizational capacity to do so and the accountability system to monitor progress, although the accountability could be an internal one rather than an external one. Elmore (2003) also talks about the importance of internal accountability, meaning schools need to take responsibility and develop a plan for improvement. Buy-in at the school level is critical for this improvement strategy to be effective.

Recently, discussions have emphasized the importance of cooperation between state- and district-level policymakers and educators. That is, when a policymaker states that a certain level of performance is expected, there also needs to be information on how the educator will be supported in their efforts to achieve that goal. This support works best when discussions are conducted both from the top down and from the bottom up. Likewise, when an educator is given additional resources, they need to indicate how these resources will be used to improve student learning.

Consider the definition of accountability offered by Stecher and Hanser (1992):

> *Accountability describes a relationship between two parties in which four conditions apply: first, one party expects the other to perform a service or accomplish a goal; second, the party performing the activity accepts the legitimacy of the other's expectation; third, the party performing the activity derives some benefits from the relationship; and fourth, the party for whom the activity is performed has some capacity to affect the other's benefits.*

To restate these conditions in an education context, policymakers ask educators to accomplish a goal; educators must see the value in those goals; educators must receive support from the policymakers; and policymakers must have the capacity to provide that support to the educators. Some of these supports may include extra time for lesson planning, increased access to professional development, and updated classroom materials.

In a report written as part of an evaluation of the implementation of state accountability systems, the American Institutes of Research (AIR, 2006) found that most states offer support to schools through strategic assistance to the districts. Few states had policies that supported schools directly. There were five common areas of support provided by the state for the districts:

1. Needs assessment and planning (revisions of district plans, data analysis tools to identify where to target resources)
2. Data analysis (state provides the analysis or tools to help districts manage large data sets)
3. Capacity building (focusing on enhancing districts' PD, leadership, or curricular activities)
4. Resource allocation (aligning budgets with priorities, providing tools to assist districts in managing and targeting funds)
5. Progress monitoring (conducting district audits and onsite reviews)

In addition to these concrete supports, some districts require schools to develop internal accountability plans that will help them raise student performance. Kirst (1990) wrote about the required supports for an accountability system to succeed in his report to the federal government. He discusses the assumption that implementing high standards, assessing students on those standards, and providing consequences for success and failure will necessarily result in higher student achievement. He argues that two necessary components of effective accountability are capacity and shared accountability. He cautions policymakers against assuming that schools have the capacity to improve and lack only the standards, assessment tools, and incentives to do so. They need to understand schools and the conditions that lead to continuous improvement and ensure that schools know what to do with the information they receive. Policymakers should develop policies that foster the leadership, collaboration, and skills that are essential to school improvement and to accountability. They also need to give the schools the time and resources they need to strengthen these areas.

Example: Massachusetts:

The Commonwealth of Massachusetts is experimenting with a pilot school option for schools that are consistently underperforming. The Commonwealth Pilot School option is patterned largely after the Pilot School model created within the Boston Public Schools, which were opened in 1995 to promote increased choice options within the school district. Pilot Schools are unique in the nation in that, by virtue of an innovative teacher union contract, they have autonomy over budget, staffing, governance, curriculum/assessment, and the school calendar to provide increased flexibility to organize schools and staffing to best meet students' needs, while operating within the economies of scale of a larger public school district. Pilot Schools were explicitly created to be models of educational innovation and to serve as research and development sites for effective urban public schools.

The Commonwealth Pilot School model is framed around the following principles:

- Provide maximum autonomy over resources, in exchange for increased accountability for student results
- Ensure buy-in and ownership of the Commonwealth Pilot School model by the school community
- Ensure that the right conditions are in place for each school to be successful
- Closely document the progress and process of each school, so that there is ample data and feedback to use in mid-course correction and improvement

Pilot Schools are subject to state and federal laws but are exempt from district policies and mandates. Pilot Schools have the following defining characteristics that enable them to be more innovative:

- Autonomous. The defining principle of Pilot Schools is that if schools are provided maximum control over their resources to be more innovative, in exchange for increased accountability, student engagement and performance will improve. Pilot Schools have autonomy over budget, staffing, curriculum/assessment, governance/policies, and the schedule.
- Accountable. Every school is held to high standards of performance through a school quality review process using a set of benchmarks that articulate the criteria for high-performing schools.
- Personalized. Pilot Schools emphasize creating a nurturing school culture in which teachers attend closely to each student's learning needs.
- Vision-Driven. Every Pilot School has a vision focused on the fundamental belief that, given the right conditions for learning, every child can achieve at high levels. Autonomy enables the school community to unify around its established vision, including having the power to hire teachers and staff who support this vision.

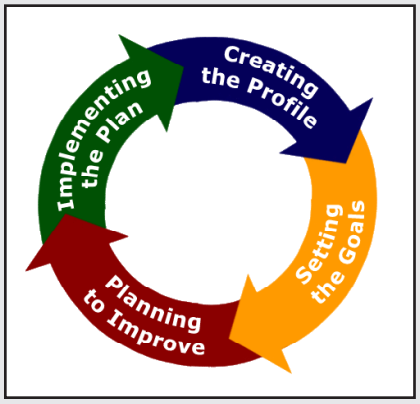
In order to convert to a pilot school, the schools are required to have the following conditions in place:

1. *Faculty Buy-In*. The school's faculty must vote by a minimum of two-thirds to convert to Pilot status.
2. *Design Team Formation and Proposal Development*. School administrators and faculty will participate with district leaders and external partners on design teams that will prepare the school for conversion into one or more Commonwealth Pilot Schools. Design teams include the principal, faculty representatives, parents, community members, and in the case of high schools, students, and may include external partners recruited to support the school's conversion into redesigned pilot schools. The design team will be responsible for crafting a proposal for the new Pilot School. The proposal must meet all of the requirements set out in the Department of Education's Commonwealth Pilot Schools Guidelines and must be presented in the application narrative form specified by the Department's pilot school request for proposals. Proposals will be submitted to the superintendent, teacher union president, and Department of Education for review and approval. Schools that do not receive approval from the Board of Education for conversion to Commonwealth Pilot Schools may be designated Chronically Underperforming schools.
3. *School Size*. Schools currently serving over 450 students must propose to downsize to a Pilot school serving a maximum of 450 students, or may propose conversion of the school into multiple small separate schools sharing a common facility.
4. *Principal Hiring and Evaluation*. The selection of a principal to lead a Commonwealth Pilot School must be mutually agreed upon by the district Superintendent and the Commissioner of Education. The school principal's annual evaluation, while the school remains in Commonwealth Pilot School status, must be the shared responsibility of the Superintendent and the school's governing board, and must be subject to oversight by the Commissioner's State Review Panel designee(s).
5. *Staff Selection and Assignment*. The principal appointed to lead a Commonwealth Pilot School has the authority to select and assign staff to positions in the school, consistent with the school's approved design, without regard to seniority. Teachers currently employed at a school being converted into Commonwealth Pilot Schools have the right to be considered first for openings at the school.
6. *Annual Work Conditions*. Teachers who work in Commonwealth Pilot Schools are exempt from teacher union contract work rules, while still receiving union salary, benefits, and accrual of seniority within the district. Teachers voluntarily choose to work at Commonwealth Pilot Schools; as a condition of employment at the school, teachers sign an "election-to-work agreement," which stipulates the work conditions for the school for the coming school year. Work conditions should be organized to maximize student learning time and provide ample time for teacher professional development and collaboration. In the first year, the terms of the teacher election-to-work agreement will be developed by the school's design team and approved by the district superintendent and teacher union leadership, consistent with the design specifications for a Commonwealth Pilot School set out in the Guidelines approved by the Board of Education. Thereafter on an annual basis the school's governing board shall revisit and revise the terms of the agreement with teacher input.
7. *School Practices*. The proposal for each Commonwealth Pilot School must provide evidence of the following practices:
    - The school administers school-wide math and English language arts assessments that are aligned with the curriculum and state frameworks, 4-6 times per year
    - The school tracks assessment results, and uses this data to inform curriculum, instruction, and individual student support
    - The school schedule provides ample time on a daily and weekly basis for instruction and individualized student support in math and English language arts (for students who are not yet proficient, the recommended amount of time is 90 minutes per subject per day)
    - The school provides before- and after-school academic support opportunities

(e.g., tutoring, learning center, peer study groups, homework help) for students who need it
    - The school has full-time faculty coaches for both English language arts and math, who are responsible for providing faculty with classroom observation and feedback on the quality of curriculum delivery, instructional practice, and data use
    - School administrators periodically evaluate faculty, including evaluation of content knowledge, tied in part to evidence of effective instruction leading to student learning, and commitment to the school's culture and pilot school model
    - The weekly and annual faculty work schedule provides ample time for regular interdisciplinary team, department, and full faculty meetings to discuss instructional practice, curriculum, assessment, school-wide improvement, and individual student progress. No less than one hour per week and no fewer than five days per year (or the hourly equivalent) should be devoted to collaborative work and professional development dedicated to instructional improvement.
8. *Governance*. Each Commonwealth Pilot School will have a governing board that includes representatives of the school's administration, faculty, parents, community members, and in the case of high schools, students, and external partners committed to achievement of the school's mission. This board is responsible for setting the vision, annual budget approval, participating in the annual evaluation of the principal (with the superintendent having final authority), and approving the annual work conditions for faculty. Each governing board will establish by-laws for meeting, membership, decisions, and communication. The initial membership of a Commonwealth Pilot School's governing board and initial appointments to board leadership roles will be mutually agreed upon by the superintendent and teacher union president.
9. *Compliance with State and Federal Laws*. Every approved Commonwealth Pilot School must comply with all federal and state laws and regulations and court orders, including those relating to diversity of students and teachers, Special Education, and English Language Learners. The local superintendent and school committee retain responsibility to ensure the quality of education offered, protect the rights and interests of students and staff, and oversee the expenditure of public funds at each Commonwealth Pilot School.
10. *Commonwealth Pilot Schools Network*. Commonwealth Pilot School administrators and faculty will be part of a statewide network of schools working together to improve student learning. The network will provide professional development opportunities and supportive activities in conjunction with the Department of Education's Commonwealth Pilot School initiative partners.

In 2007, four schools optioned to become pilot schools after receiving low scores on the MCAS and being designated as "under-performing" by the state accountability system.

Example: Nebraska



Nebraska uses a continuous improvement process to support its districts' efforts to continually improvement student performance. There are four phases to this process:

1. Creating the Profile, which includes engaging stakeholders, collecting and analyzing data, and reflecting on the profile;
2. Setting the Goals, which includes determining strengths, identifying priorities, and establishing areas of need;
3. Planning to Improve, which includes determining the action plan, establishing a timeline and locating resources; and
4. Implementing the Plan, which includes monitoring progress and determining the effectiveness of the interventions.

This four phase plan helps the district improvement process by:

- Building on existing efforts to improve student achievement
- Encouraging a continuous process
- Integrating activities and programs
- Incorporating researched practices (Effective Schools)
- Using the rubric for school improvement developed by Nebraskans for Nebraskans
- Identifying strategies for targeting areas of low-performance

Example: Maine

For Title I schools that have been identified as being a Continuous Improvement Priority School (CIPS), state officials send in a school improvement team. Two distinguished educators are part of that team, assuring teachers are assisted by other educators. Currently, the distinguished educators are former building administrators. In addition, the state department of education in Maine has developed in online data tool called "I Analyze." The schools use this tool to sort, chart and analyze reading and math data in all groups. It also gives them the power to print out reports for use with individual teachers, teams of teachers, entire staff, administrators or school board members.

With schools entering a year one CIPS designation, the school improvement team primarily focuses on short term strategies, such as Test Preparation (e.g., curriculum alignment, appropriate use of accommodations, and allowing students to take a practice test), Test Conditions (such as allowing students to take the test in their own classroom with their own teachers and establishing appropriate schedules and breaks), and Motivation (such as communicating with parents, coaching the students appropriately, and providing small rewards like certificates, pencils, or snacks). In year 2, the focus moves to very targeted and focused teacher professional development, which is where the distinguished educators are particularly helpful. In addition, the team often brings in specific training consultants to work with individual teachers in the classroom and then provides 3–5 days of professional development in the summer for the entire staff including educational technicians, special educators, and administrators. Some specific training examples include Connected Math Training, Everyday Math Training, Literacy Audits, Formative Assessment Training, and Data Team Training. The summer training leads to additional training the following school year in the classroom with teachers on specific teaching strategies they have learned and how to better implement that strategy to increase student achievement.

State officials in Maine have also been developing a "Readiness" rubric that school can use to help determine their capacity to do school improvement work. This rubric can be found in Appendix D of this report.

Example: Ohio

The Ohio Department of Education (ODE) is piloting the use of diagnostic tools as part of their district and school improvement process. ODE began with a simple diagram (a circle) that represents the improvement process. There are four defined steps in the process: Identify Needs; Develop a Focused Plan; Implement the Focused Plan; Monitor/Evaluation the Implementation/Success of the Focused Plan. The circle is used as a reference point for communicating with internal/external participants and clarifies the major steps leading toward improvement. Additionally, as additional resources are identified, the circle is used to organize the resources and help users understand in which step of the process the resources would be most useful or appropriate.

The assumptions that guide the improvement process and development of diagnostic tools are that all districts and schools need improvement, but some need more improvement than others, and some need more intensive/frequent help than others. Therefore, this process allows for differentiated services and resources based on need. The tools are developed for all districts to use, but in some cases, the state (or its representative) will participate as an oversight to or partner in the use of the tool. For example, the diagnostic tool would be used by a state appointed team in districts/schools the State determines to be

high priority districts/schools. The diagnostic tool will also be available for any other district in the state to use to learn more about itself. It could choose to conduct the diagnostic review and could work with an Educational Service Center (ESC) to train district staff to conduct the review; train and conduct a peer review team from surrounding districts; or conduct the review with trained ESC staff.

Currently, there are two tools that are combined to create the needs assessment (step 1 of the process): the diagnostic tool and the decision framework. The diagnostic tool is used to audit behaviors of districts/schools. The focus is to create data in areas where data is generally not available to the state and often the district—referred to as "soft" data. For example, a school might have data on teacher quality, but nothing that indicates if teachers use assessment data to guide classroom instruction or if classroom instruction is aligned to the state standards. The diagnostic tool then seeks to develop as set of data that can be used to help the state, district, and school understand district/school behaviors. The diagnostic process is conducted by a trained team that uses observations, individual and group interviews and document review to make qualitative determinations in the areas of:

- Alignment with standards
- Instructional practices
- Environment/Climate
- Leadership
- Professional development
- Data driven decisions

The results of the review are translated into a numerical grade that helps inform the strengths and weaknesses of the district/schools.

This "soft" data along with other data is then moved through the decision framework (the second tool). The decision framework sorts/organizes the soft and hard data to assist in identifying the major problems of the district/schools, the most probable causes of those problems and the resource management practices that might be contributing influences, such as money, time, or personnel. The process takes pieces of the data and forces the district via "essential" questions and required responses to make informed determinations related to the severity of the issue or the contribution of the issue to the problem. The content areas are addressed first, forcing the district to determine which content area or areas needs the most attention. Of those areas rising to the top, the district/building team further reviews data in the following sub-categories: teacher quality; professional development; and curriculum, instruction, and assessment. This part of the process includes sub-scale and disaggregated data. Next the data related to expectations and conditions is examined in the areas of leadership, climate and parent/community. Finally, districts are asked to respond to essential questions related to their resource management. The result of this process is a very clear Needs Assessment that should be easily translated into the focused goals, focused strategies and action steps.

An issue in some of the poor performing districts and schools is teacher retention. Elmore (2003) emphasizes the importance of creating incentives to stabilize the population of teachers and administrators in low-performing schools. These incentives may be in the form of financial incentives, such as loan forgiveness or extra compensation; professional development, including improved access to courses and advanced degree programs; or improved working conditions, such as reduced teaching loads, improved physical infrastructure, and better materials. For example, Ohio has developed a series of strategies for improving teacher retention, particularly in the high-need schools, that include the following:

- Provide mentor training to experienced teachers to assist them in mentoring entry-year teachers
- Provide incentives to highly qualified, experienced teachers who teach in high-need schools
- Target support to National Board certification candidates who teach in regions with high numbers of high-need schools and provide incentives so they stay
- Provide an opportunity for "high value added" teachers in high-need schools to apply for incentives and rewards to encourage them to stay

- Place subject specialists/coaches in high-need schools; Provide targeted coaching to teachers who teach reading and mathematics
- Make a major investment in preparing principals for the unique challenges of high-need schools (e.g., how to promote student learning, how to hire strong teachers)
- Keep good teachers where they are needed most by providing tools and resources for improving the working conditions in high-need schools
- Provide monetary incentives to encourage highly qualified teachers who teach in high-need schools to stay
- Ensure that state testing policies and systems of rewards and sanctions do not inadvertently drive teachers and principals away from high-need schools

References and Suggested Readings for This Chapter

American Institutes for Research. (2006). *Summary of state strategies for districts identified for improvement under NCLB*. Sacramento, CA: California Comprehensive Center.

Elmore, R. (2003). *Knowing the Right Thing to Do: School Improvement and Performance Based Accountability*, Washington, DC: NGA Center for Best Practices

Kirst, Michael W. (1990). *Accountability: Implications for State and Local Policymakers*. Washington, DC: Information Services, Office of Educational Research and Improvement, U.S. Department of Education.

Newmann, F., King, M. & Rigdon, M. (1997). Accountability and School Performance: Implications from Restructuring Schools. *Harvard Education Review*, 61(1), pp. 41–69.

Stecher, B. & Hanser, L. (1992). Local accountability in vocational education: A theoretical model and its limitations in practice. MDS-291. Santa Monica, CA: RAND.

Checklist for State Accountability Systems: System Monitoring, Evaluation, and Improvement

- *How will the accountability system be monitored and evaluated?*
  - □ Plan for annual monitoring and evaluation of the system.
  - □ Consider state capacity for monitoring.
- *How can the intended and unintended effects of the accountability system be evaluated?*
  - □ Plan and implement longitudinal studies evaluating the effects of the accountability program.
  - □ Determine whether unintended effects of the accountability system are positive or negative and minimize unintended negative effects.
  - □ Conduct ongoing evaluations of the validity of the indicators.
  - □ Consider using an external evaluator to independently review the system and provide confirming (or disconfirming) evidence to supplement the internal evaluation
- *How will the results of this monitoring and evaluation be used to improve the system?*
  - □ Avoid frequent or drastic changes to the accountability system as they reduce credibility and public support.
- *How will the system design incorporate the need for revisions over time?*
  - □ Develop steps to continuously evaluate and improve the effectiveness of the accountability system itself.
  - □ Take steps to evaluate and improve the theory of action as well as any of the components.
  - □ Ensure the system is sufficiently flexible to support additional indicators as new data become available.

The last core concept addresses the need to monitor, evaluate, and improve the accountability system. The core elements in this section focus on the mechanism for continually analyzing the adjusting the model appropriately. They highlight the need to develop an evaluation plan and then use that information to make improvements. This evaluation should also answer questions regarding the effectiveness of various rewards and sanctions as well as other intervention or support strategies.

Researchers seem to agree that an accountability system should include a mechanism for continuously monitoring and evaluating the effects of the system and devise strategies to adapt and improve the system in response to new information. A key question is how the system design will incorporate the need for revisions over time. State- and district-level policymakers need to have a predetermined plan of how they will manage deficiencies uncovered by the accountability system and how their solutions will feed back into the system itself. As mentioned in a previous section, policymakers should monitor how schools and educators respond to sanctions and rewards.

In addition, individual components of the accountability system should be monitored. The quality of the accountability systems is only as good as the quality of the indicators. As mentioned in the chapter on indicators, a quality assurance plan should be developed to examine the accuracy and validity of each indicator used in the accountability system. For example, Taylor, Beaudoin, and Goldschmidt (2007) recommend internal auditing, data collector verifications, rate assignments, trend analysis, risk assessments, external audits, on-site monitoring, and data collector trainings as steps to implement during data collection and reporting.

Baker, Linn, Herman, & Koretz (2002) list evaluation as one of the five main areas of accountability. They assert:

> Longitudinal studies should be planned, implemented, and reported evaluating effects of the accountability program. Minimally, questions should determine the degree to which the system:
>
> a. builds capacity of staff;
> b. affects resource allocation;
> c. supports high-quality instruction;
> d. promotes student equity access to education;
> e. minimizes corruption;
> f. affects teacher quality, recruitment, and retention; and
> g. produces unanticipated outcomes.

They go on to say that the evaluation must determine the degree to which the intended benefits are realized and the costs in terms of unintended negative consequences are minimized. For example items (a) through (d) above reflect intended positive consequences the realization of which is the focus of evaluation. Items (e) and (g) emphasize the needed evaluation of plausible unintended negative consequences. Item (f) requires the evaluation of both intended positive and unintended negative influences of the accountability system.

Gong (2002) also lists evaluation and monitoring as a key design principal for accountability systems. He recommends asking questions such as:

- Is the system complete?
- Can the system be improved?
- Is the system having the desired effects?
- Is the system producing any undesired effects?
- Have assumptions or circumstances changed to an extent that the system should change?

Other sources for finding information on monitoring and evaluation include the *Program Evaluation Standards* (JCSEE, 2004). Nebraska, Kentucky, and Delaware have all implemented evaluation policies and developed analyses and quality assurance techniques to monitor an accountability system in its early stages.

Example: Nebraska

Because Nebraska has a locally-based accountability system, they have one of the most extensive monitoring systems in place. Their monitoring system includes peer review of the local systems, benchmark checks, external auditing, and moderation strategies. The state provides technical assistance to its districts, particu-

larly as they address needs of students with disabilities, English language learners, migrant students, and homeless students. The state is currently implementing a student information system across the state that is expected to further assist efforts in monitoring and supporting local school districts.

Example: Delaware

Delaware examines its accountability system annually through its accountability manual. Beginning in SY 2002-2003, the Delaware Department of Education published this document to record design changes, score results, and quality assurance efforts associated with the State's accountability system. The agency conceptualized a technical document having detailed information about how: (a) the accountability system processed data into scores and ratings, (b) the business rules and design logic were implemented, and (c) the quality assurance mechanisms ensured accurate and credible results. During SY 2003-2004, the first technical manual (*Technical and Operational Manual-2004*) focused on several accountability characteristics. These characteristics were expanded, refined, and strengthened upon each subsequent publication; however, the document's primary function remains as:

1. A document used to understand how Delaware's accountability system evolved into its current configuration;
2. A map detailing the alignment between statutory requirements, business rules, and score production sequences;
3. A foundation for examining ways to improve the accountability system, and;
4. An examination of quality assurance practices used within the agency.

This document provides a historical context for the reader, along with the state's theory of action. The technical manual documents in meticulous detail the business rules, design logic, and operational sequences used to produce accountability scores and ratings. The results for the current year are displayed in the Implementation section. Here, accountability results are examined using a series of analytical solutions. The Quality Assurance section reflects the Department's continued improvement efforts.

The prior version (2005) of this manual was completely renovated based upon the Department's internal efforts and the recent publication of *Validity Threats: Detection and Control Procedures for Local and State Officials* (CCSSO, 2006). The quality assurance section of the current report documents those efforts used to build and strengthen the State's error detection, control, and improvement mechanisms.

Example: Maine

Maine is a locally-controlled state. Although the NCLB assessments are developed at the state level, each school is responsible for developing a comprehensive education plan. Included in this plan are locally-developed assessment systems for all subjects in their system of Learning Results. These local assessment systems will combine classroom, school, district, and state assessments and determine how to include and weigh multiple measures. As part of this plan, schools are encouraged to develop internal monitoring and evaluation systems. As stated in chapter 125 of their state statute: "School administrative units shall demonstrate how school and student assessment data are used to evaluate, develop, and improve curriculum, instruction, and assessment, as specified in the Comprehensive Education Plan." Accompanying this directive is one from chapter 127 of the same document: "The school board shall annually review and publish school and school administrative unit results on the local assessment system, and, if required based on these results, shall adjust the Comprehensive Education Plan developed in accordance with Me. Dept. of Ed. Reg. 125."

The state conducts a review of the local systems by content area on a four-year cycle that began in the 2003-2004 school year. The state may also review the school's Comprehensive Education Plan if the statewide assessment results indicate that a school is not meeting its annual goals. This review will include an analysis of student performance on other indicators specified in the unit's Comprehensive Education Plan, school approval requirements specified in Me. Dept. of Ed. Reg. 125, and the administrative unit's programs and services designed to help students meet the content standards of the system of Learning Results. If warranted based on this analysis, the state Commissioner will assign a team with expertise in the

areas of need identified in the analysis to assist in planning for improved student performance, for a period of at least one year.

Currently, state officials only intervene when a school has been designated a Continuous Improvement Priority School (CIPS). However, state officials have recognized a need for greater state oversight, so this legislation is being revised as of the summer of 2007.

Conclusion

The importance of continually monitoring, evaluating and improving educational accountability cannot be understated. An accountability system should be considered a living entity that needs to be able to incorporate new indicators, changes in how indicators are combined and used to make decisions, and improvements in reporting and communication. The checklist in Appendix C is intended to be used as a tool for evaluating current systems with an eye towards enhancing the validity of the system. All of the core concepts discussed in previous chapters are incorporated into this final step of monitoring, evaluating, and improvement and all should be revisited throughout the life of any accountability system.

References and Suggested Readings for This Chapter

ASR SCASS. (2006). *Validity Threats: Detection and control practices for state and local education officials*. Washington, DC: Council of Chief State School Officers

Baker, E., Linn, R., Herman, J., & Koretz, D. (2002). *Standards for educational accountability systems* [Policy Brief 5]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing

Gong, B. & ASR SCASS. (2002). *Designing school accountability systems: Toward a framework and process*. Washington, DC: Council of Chief State School Officers.

Joint Committee on Standards for Education Evaluation. (2004). *The Program Evaluation Standards. How To Assess Evaluations of Educational Programs*. (2nd Ed.). Thousand Oaks, CA: Sage Publications.

LeFloch, K., Taylor, J., & Thomsen, K. (2005). *The implication of NCLB accountability for comprehensive school reform*. Washington, DC: American Institutes for Research.

Stecher, B. & Hanser, L. (1992). Local accountability in vocational education: A theoretical model and its limitations in practice. MDS-291. Santa Monica, CA: RAND.

Taylor, R., Beaudoin, J. & Goldschmidt, P. (2007). *Quality assurance practices associated with producing cohort graduation rates*. Washington, DC: Council of Chief State School Officers.

APPENDIX A: ANNOTATED BIBLIOGRAPHY

American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

This resource lists the standards professionals should use in designing and evaluating assessments. The standards are organized into three main areas: test construction, evaluation and documentation; fairness in testing; and testing applications.

American Evaluation Association (2002). *Public statement: High stakes testing in preK-12 education*. Fairhaven: MA: American Evaluation Association. Downloaded October 8, 2007 from http://www.eval.org/Publications/GuidingPrinciples.asp.

This public statement describes concerns about relying on high-stakes tests to make decisions about students, teachers and schools without fully studying the validity of these tests. It lists numerous violations of professional standards in using high stakes testing programs for evaluative purposes. It also lists expectations for improved evaluation practice.

American Evaluation Association (2004). *Guiding principles for evaluators*. Fairhaven: MA: American Evaluation Association. Downloaded October 8, 2007 from http://www.eval.org/hst3.htm.

This document was originally published in 1994 and revised in 2004. It lists a set of five general principles to guide evaluators, and then provides a set of statements under each principle to provide more specific direction for evaluators.

American Evaluation Association (2006). *Public statement: Educational Accountability*. Fairhaven: MA: American Evaluation Association. Downloaded October 8, 2007 from http://www.eval.org/edac.statement.asp.

This public statement describes the relationship between good evaluation and good accountability and applies AEA's guiding principles for evaluators to the field of accountability. It identifies three main concerns with current educational accountability systems and provides suggestions for creating more valid accountability systems.

American Institutes for Research. (2006). *Summary of state strategies for districts identified for improvement under NCLB*. Sacramento, CA: California Comprehensive Center.

This research report summarizes the supports and interventions states are using in districts identified as needing improvement or in corrective action.

Ananda, S. & Rabinowitz, S. (2001). *Building a workable accountability system*. San Francisco: WestEd. Downloaded February 20, 2007 from www.WestEd.org/online_pubs/kn-01-03.pdf.

This brief for policymakers and educators identifies the questions that must be answered in designing or evaluating an accountability system. It summarizes the issues, provides options, and also identifies potential problems.

ASR SCASS. (2006). *Validity Threats: Detection and control practices for state and local education officials*. Washington, DC: Council of Chief State School Officers

This report examines the validity threats to any accountability system and recommends quality control activities to counter each threat. It examines the three central components of accountability systems: inputs, processes, and outputs, and explores the vulnerabilities and validity checks required at each stage.

Bailey, A., Winter, P., & CAS SCASS. (2002). *Incorporating federal requirements into state accountability plans*. Washington, DC: Council of Chief State School Officers.

This CCSSO report was developed to provide guidance to state and educational leaders on developing comprehensive accountability systems for students at all levels. It includes in-depth case studies of four states and discusses ways to integrate state and federal accountability systems.

Baker, E. & Linn, R. (2002). *Validity issues for accountability systems* [Technical Report 585]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing

This report discusses the conditions that must be present for accountability systems to result in building on strengths and reducing weakness. It considers the validity of the use of the assessment in the accountability system and discusses alignment from test items to test specifications to content standards to reporting domain. Finally discusses shared responsibility for results in any accountability system.

Baker, E., Linn, R., Herman, J., & Koretz, D. (2002). *Standards for educational accountability systems* [Policy Brief 5]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing

This document puts forth 22 standards for educational accountability systems, divided into five domains: System components, testing, stakes, public reporting, and evaluation. These standards are intended to provide guidance to states and districts in developing or reviewing their own accountability systems. These standards represent models of practice from three perspectives: research knowledge, practical experience, and ethical considerations.

Barton, P. E. (2006). "Failing" or "Succeeding" schools: How can we tell? Washington, DC: American Federation of Teachers. Retrieved February 20, 2007, from http:// www.aft.org/pubs-reports/downloads/teachers/101006FailingorSucceeding.pdf.

This report focuses on standards-based reform, discussing how tests are the driver of most accountability systems. It focuses on requirements of these tests, such as transparency and alignment, for them to be an effective component of accountability.

Carlson, D. (2002). An analysis of some approaches to combining multiple measures and making decisions in state accountability systems. In W. J. Erpenbach et al., *Incorporating multiple measures of student performance into state accountability systems—A compendium of resources* (pp. 285-297). Washington, DC: Council of Chief State School Officers.

This report presents a framework for considering effective schools. It presents an analysis of six state accountability systems and how they combine the various indicators of quality.

Carlson, D. (2006). *Focusing state educational accountability systems: Four methods of judging school quality and progress*. Dover, NH: National Center for the Improvement of Educational Assessment.

This paper focuses the discussion on how to answer two fundamental questions in accountability: (1) How good is this school? (2) Is it getting better? It compares status measures to change measures (both improvement and growth), providing a framework of four quadrants to help think through what measures make sense under the assumptions of different accountability systems.

Carnoy, M., Elmore, R., & Siskin, L., eds. (2003). *The New Accountability: High Schools and High Stakes Testing*. New York: Routledge.

This edited book contains eight chapters discussing issues of assessment and accountability in high schools. It focuses on various aspects of a study conducted by the authors that exampled the impact of state accountability systems on the internal accountability of high schools in four states.

Chester, M. (2005). Making valid and consistent inferences about school effectiveness from multiple measures. *Educational Measurement: Issues and Practice, 24*(4), 40-52.

This article reports results from a study exploring the use of multiple measures to enhance the validity and reliability of inferences about school and district effectiveness. Data from Ohio are used to explore ways to combine data from multiple measures to make both individual and system-level decisions.

Cross, R., Rebarber, T., & Torres, J., Eds. (2004). *Grading the system: The guide to state standards, tests, and accountability policies*. Washington, DC: Thomas B. Fordham Foundation.

The publication starts with the premise that standards-based education has three components: standards, assessments, and the accountability system or consequences. The majority of the report provides the results of a study that examined 30 states on six measures: the quality of the reading and math standards, the content of the tests, the alignments between standards and tests, the tests' rigor (based on how high the cut score was set), their technical reliability, and the accountability policies. The report both analyzes and rates the 30 states, providing state profiles and grades.

Darling-Hammond, L. (2006). *Standards, assessments, and educational policy: In pursuit of genuine accountability*. Princeton, NJ: ETS

Part of the ETS Angoff Lecture Series, this paper discusses how to use assessments appropriately in a standards-based accountability system. It concludes with a list of principals for effectively using assessments for school improvement.

Elmore, R. (2003). *Knowing the Right Thing to Do: School Improvement and Performance Based Accountability*, Washington, DC: NGA Center for Best Practices

This paper explores the issues of how an accountability system can push schools to get better. It focuses on two low performing schools to determine the necessary elements in accountability for improving the lowest performing schools.

Elmore, R. (2004). The Problem of Stakes in Performance-Based Accountability Systems. In Fuhrman, S. & Elmore, R. (eds.) *Redesigning Accountability Systems for Education*. New York, NY: Teachers College Press.

This final chapter in a book about understanding and designing accountability systems summarizes many of the discussions regarding school improvement models, validity issues, inclusion, effects on students, teachers, and schools, and to lay out some accountability design principles.

Erpenbach, W. (2002). *Incorporating multiple measures of student performance into state accountability systems—A compendium of resources*. Washington, DC: Council of Chief State School Officers.

This compendium is on a DVD and includes a report, supporting documentation, and other resources. The report synthesizes previous research and resources to develop a list of required elements in accountability systems. It focuses on the need for multiple measures and how to incorporate them effectively into an accountability system.

Erpenbach, W., Forte Fast, E., & Potts, A. (2003). *Statewide educational accountability under NCLB*. Washington, DC: Council of Chief State School Officers.

This CCSSO report summarizes state accountability workbooks prepared under NCLB. It discusses some central issues and documents the U.S. Department of Education's reviews and responses to these issues.

Forte Fast, E. & ASR SCASS. (2003). *A guide to effective accountability reporting*. Washington, DC: Council of Chief State School Officers.

This monograph by CCSSO is intended to serve as a resource to state department of education staff who are responsible for producing state, district, or school report cards for accountability purposes. It outlines a process from identifying reporting needs to determining the content to designing the format.

Forte Fast, E. & Erpenbach, W. (2004). *Revisiting statewide educational accountability under NCLB*. Washington, DC: Council of Chief State School Officers

This CCSSO report contains a summary of state requests for amendments to their NCLB accountability plans in 2003-04 and their approval status. Note that similar reports have been prepared every year since 2004, and this is representative of them.

Forte Fast, E. & Hebbler, S. (2004). A framework for examining validity in state accountability systems. Washington, DC: Council of Chief State School Officers

This CCSSO report lays out a framework under which one can examine the validity of a state accountability system, focusing on four core concepts: Goals, consequences, decision rules, and performance indicators.

Fuhrman, S. (1999, January). The new accountability. *CPRE Policy Briefs*, RB-27. Philadelphia, PA: The Consortium for Policy Research in Education.

This policy brief discusses the changes over time in accountability with a focus on school accreditation. It also lists issues associated with designing accountability system and provides some required elements for these systems to be effective.

Fuhrman, S. (2003, September). Redesigning accountability systems for education. *CPRE Policy Briefs*, RB-38. Philadelphia, PA: The Consortium for Policy Research in Education.

This policy brief is intended as a summary of the book Redesigning Accountability Systems for Education by Fuhrman and Elmore (2004). It focuses on four questions regarding the validity, fairness, effects, and future improvement of accountability systems.

Fuhrman, S. & Elmore, R., (2004). *Redesigning Accountability Systems for Education*. New York, NY: Teachers College Press.

This book is about understanding and designing accountability systems. This edited volume brings together experts in the field to discuss school improvement models, validity issues, inclusion, effects on students, teachers, and schools, and to lay out some accountability design principles.

Glass, Gene V. (1978) "Standards and Criteria." *Journal of Educational Measurement, 15*, pp 237-261.

This seminal article focuses on the evolution of criterion-referenced tests and performance standards. It describes a variety of methods for determining minimal competency and setting a cut score.

Goertz, Margaret. (2001). The Federal Role in an Era of Standards-Based Reform. in *The Future of the Federal Role in Elementary and Secondary Education: A Collection of Papers*. Washington, DC: Center for Education Policy.

This article traces the progression of various federal legislation, guidance, and regulations for Title I and Title III. It discusses changes in the federal role in accountability systems and provides guidance for future reauthorization of ESEA (prior to NCLB).

Gong, B. & ASR SCASS. (2002). *Designing school accountability systems: Toward a framework and process*. Washington, DC: Council of Chief State School Officers.

This CCSSO report is meant to assist states in making decisions about designs for accountability systems. It lays out ten central questions that must be answered in any accountability design, and then explores possible directions states can go in answering the questions.

Goodwin, B., Englert, K. & Cicchinelli, L. (2003). *Comprehensive accountability systems: A framework for evaluation*. Aurora, CO: Mid-continent Research for Education and Learning.

This report offers a set of recommendations and guidelines for accountability systems. Three models for conceptualizing accountability systems are presented: a purpose-oriented accountability system, an input-process-output accountability system, and an

issue-oriented accountability system. It also discusses questions and alternatives that appear to be missing from the current dialogue surrounding accountability systems.

Haertel, E. H. (1999). Validity arguments for high-stakes testing: In search of the evidence. *Educational Measurement: Issues and Practice, 18*(4), pp. 5-9.

In this NCME presidential address, Dr. Haertel outlines issues related to evaluating the validity of test results. He questions the premises and the motivations for conducting such analyses and encourages measurement professionals to think about developing and evaluating less common questions about validity.

Hamilton, L. & Stecher, B. (2004). Responding Effectively to Test-Based Accountability. *Phi Delta Kappan, 85*(8). pp. 578-583.

The authors address the need to make NCLB work for schools and districts as a school improvement tool. They provide suggestions for building capacity, using standards to improve instruction, using assessment results appropriately, creating effective incentives, and helping parents make effective choices.

Hamilton, L., Stecher, B., & Klein, S. (Editors.) (2002). *Making sense of test-based accountability in education*. Santa Monica, CA: RAND

This book is intended for policymakers and educators as a tool for helping them understand test-based accountability and using it effectively. Specifically, the book provides suggestions for selecting tests and using the scores by focusing on guidelines in four areas: (1) how tests are used in accountability systems; (2) how to evaluate the technical quality of tests; (3) how test-based accountability affects teacher practices; and (4) how political considerations affect the debate.

Hanushek, E. & Raymond, M. (2002). Sorting out accountability systems. In W. Evers & H. Walberg (Eds.) *School accountability* (pp. 75-104). Palo Alto, CA: Stanford University, Hoover Press.

This chapter reflects on the history of educational accountability and compares that to current (pre-NCLB) systems. It examines the necessary components in an effective system and discusses issues in implementation.

Herman, J. & Haertel, E. (Eds.) 2005. Uses and misuses of data for educational accountability and improvement. 104th yearbook of the National Society for the Study of Education, Part 2. Malden, MA: Blackwell Publishing.

This 103rd yearbook of the National Society for the Study of Education was written three years after the passage of No Child Left behind. It discusses the role of assessment in accountability and educational improvement. Various chapters deal with foundational issues, the policy and values underlying current accountability testing, technical concerns such as validity and alignment, and issues of fairness and consequences.

Hill, R. & DePascale, C. (2003). Reliability of No Child Left Behind accountability designs. *Educational Measurement: Issues and Practice, 22*(3), 12-20.

This article discusses the competing concerns of validity and reliability in accountability models, focusing specifically on NCLB. It shows that it is difficult to design an accountability system that is both highly reliable and highly valid, particularly if one follows NCLB to the letter of the law.

Joint Committee on Standards for Education Evaluation. (2004). *The Program Evaluation Standards. How To Assess Evaluations of Educational Programs. (2nd Ed.)*. Thousand Oaks, CA: Sage Publications.

This report issued a set of standards to be used in assessing the validity and completeness of program evaluation. A total of 28 standards were developed, including Proprietary Standards, Utility Standards, Feasibility Standards, and Accuracy Standards.

Joint Committee on Standards for Education Evaluation. (2002). *The Student Evaluation Standards. How To Improve Evaluations of Students*. Thousand Oaks, CA: Corwin Press.

This report presents 28 standards for designing, implementing, assessing and improving student evaluation. Each of the 28 standards has been placed in 1 of 4 essential categories: Propriety, Utility, Feasibility, and Accuracy. The primary focus of these standards is to promote sound, credible, and accurate evaluations that foster student learning and development at the classroom level. These standards are intended for teachers and others who evaluate students as well as those who use and are affected by student evaluations.

Jones, K. (2004). A balanced school accountability model: An alternative to high-stakes testing. *Phi Delta Kappan, 85*(8). pp. 584-590.

This article proposes a balanced and comprehensive accountability model that relies on indicators beyond assessment. It proposes that schools should be accountable for more than student learning and provides a framework for evaluating the health of an organization that is currently used in the business world.

Kelley, C. (2002) Financial incentives in state accountability systems: Performance pay for teachers. Madison, WI: Consortium for Policy Research in Education.

This paper describes some of the research linking monetary incentives for teachers to improved student performance. It describes conditions that must be in place for school-based performance awards to be effective.

Kirst, Michael W. (1990). *Accountability: Implications for State and Local Policymakers*. Washington, DC: Information Services, Office of Educational Research and Improvement, U.S. Department of Education.

This report discusses six broad approaches to accountability: (1) accountability through performance reporting; (2) accountability through monitoring and compliance with standards/regulations, (3) accountability through incentive systems; (4) accountability through reliance on the market; (5) accountability through changing the locus of authority or control of schools; and (6) accountability through changing professional roles. It contends that state or local governments must use several of these approaches simultaneously.

LeFloch, K., Taylor, J., & Thomsen, K. (2005). *The implication of NLCB accountability for comprehensive school reform*. Washington, DC: American Institutes for Research.

This paper summarizes findings of a national comprehensive school reform study, using longitudinal data to examine the potential conflict between comprehensive school reform and No Child Left Behind. The authors address the question of how states pursuing comprehensive state reform can implement NCLB accountability mandates in a coherent approach to school improvement.

Levin, H. (1974). A Conceptual Framework for Accountability in Education. School Review 82(3), pp. 363-391.

This paper provides an early framework for considering educational accountability. It includes one of the first definitions for accountability and provides a list of key elements to consider.

Linn, R. (2001). The design and evaluation of educational assessment and accountability systems. [Technical Report 539]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

This technical report discusses variances in state accountability systems prior to the passage of NCLB. It discusses differences in status and improvement models, focuses on the inclusion of students from different socioeconomic backgrounds, and evaluates the degree to which performance reports support inferences about student learning.

Linn, R. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher, 32*(7), pp. 3-13.

The AERA presidential address discusses key elements of accountability systems. Dr. Linn argues for shared responsibility in creating effective schools and discusses designs for effective accountability systems based on research and practice. These designs are then compared to NCLB.

Linn, R. (2005a). *Test-based educational accountability in the era of No Child Left Behind*. [Technical Report 651]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

This technical report discusses the mixed messages sent by NCLB and state accountability systems. It points to features of NCLB accountability that contribute to these mixed messages including multiple hurdles, status versus change, and the definition of the performance goals. Suggestions are made for improving NCLB.

Linn, R. (2005b). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *Educational Policy Analysis Archives, 13*(33). Retrieved January 15, 2007 from http://epaa.asu.edu/epaa/v13n33.

This article details the tension between NCLB and state accountability systems. Kentucky, Florida, and Colorado are given as examples of states in which many schools receive mixed messages about their effectiveness from the two different systems. The article describes differences in the two types of systems, explains why these conflicts

occur, and provides guidance on what accountability systems should be measuring and reporting.

Linn, R. (2006). *Educational accountability systems*. [Technical Report 687]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

This technical report describes a variety of test-based accountability systems developed by states under NCLB and evaluates their strengths and weakness. It also explores the inferences of school effectiveness implied by the results of the accountability system, arguing that these systems should serve as a source of information rather than the sole determinant of sanctions.

Linn, R., Baker, B., & Betebenner, D. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher, 31*(6), pp. 3-16.

This article discusses the challenges for states in developing assessment systems that meet the requirements of NCLB. The differences in states content and performance standards are discussed and compared to state NAEP. The authors provide suggestions for using NAEP to level the playing field and to bring some consistency in meaning of the performance levels across states.

Louis, K., Febey, K., & Schroeder, R. (2005). State-mandated accountability in high schools: Teachers' interpretations of a new era. *Educational Evaluation and Policy Analysis, 27*(2), pp. 177–204.

This article examines the impact of accountability policies at the high school level. It explores the arguments that testing undermines good teaching or that it stimulates improvement. The authors begin with the assumption that it is important to explore implementers' cognitive perspectives in order to understand a policy's effects. They examine teachers' responses to accountability in three high schools in states with diverse histories of accountability legislation, but in districts with well-established local accountability policies.

Marion, S., White, C., Carlson, D., Erpenbach, W, Rabinowitz, S., & Sheinker, J. (2002). *Making valid and reliable decisions in determining adequate yearly progress*. Washington, DC: Council of Chief State School Officers.

This CCSSO report is intended for state policymakers involved in educational accountability. It explores key issues, decision points, decision consequences, and policy implications related to making valid and reliable decisions in the calculation of adequate yearly progress. It explores critical variables and unique technical issues in designing accountability systems under NCLB.

Mintrop, H. & Trujillo, T. (2007). *The practical relevance of accountability systems for school improvement: A descriptive analysis of California Schools*. [CSE Report 713]. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

This report presents the findings from a study comparing nine exceptionally high and low performing urban middle schools within the California accountability system looking for substantive differences between the schools.

National Research Council (1999). *Testing, teaching, and learning: A guide for states and school districts*. Elmore, R. & Rothman, R., Eds. Washington, DC: National Academy Press.

This report explores the practical implications of Title I assessment and accountability prior to NCLB. It is intended to guide policymakers as they develop and improve systems of assessment and accountability. It includes a framework for developing assessment and accountability systems intended to improve the quality of instruction in schools.

National Research Council (2001). *Knowing what students know: The science and design of educational assessment*. Pelligrino, J., Chudowsky, N., & Glaser, R. (Eds.). Washington, DC: National Academy Press.

This book written by NRC's Board on Testing and Assessment has helped reshape researchers' approach to testing with its focus on cognition. Understanding the cognitive sciences that are the underpinning to learning theory provides a framework for developing better, more informative items. It also describes innovative approaches in technology for learning and assessment.

National Research Council (2006). *Lessons learned about testing: Ten years of work at the National Research Council*. Washington, DC: National Academy Press.

This booklet produced by the NRC's Board on Testing and Assessment summarizes some recurring themes in testing. It summarizes work by the NRC on proper uses of testing, focusing on uses, design consequences, and public understanding.

Newmann, F., King, M. & Rigdon, M. (1997). Accountability and School Performance: Implications from Restructuring Schools. *Harvard Education Review, 61*(1), pp. 41–69.

This article explores the theory that strong external accountability will result in improved student achievement. The authors examine 24 restructuring schools and highlight the importance of organizational capacity and internal accountability.

Novak, J. R. & Fuller, B. (2003, December). Penalizing diverse schools? Similar test scores, but different students, bring federal sanctions. *Policy Brief 03–4*, Berkeley, CA, University of California Berkeley: Policy Analysis for California Education, available online at http://pace.berkeley.edu.

This brief discusses some of the conflicts between state and federal accountability systems, comparing California's API with the federal AYP requirements. It highlights concerns that schools with more subgroups counted are more likely to miss their accountability targets than schools that serve a less diverse population.

O'Day, J. (2002). Complexity, accountability, and school improvement. *Harvard Educational Review, 72*(3), pp. 293-329.

This chapter looks at accountability systems where the school is the unit of accountability and discusses how these systems fit into a theoretical framework on organizational learning and adaptation. The evidence includes case studies describing and comparing accountability systems in two large central-city school districts: Chicago Public Schools and Baltimore City Schools.

Olson, L. (2006, December 6). U.S. urged to rethink NCLB "tools": Districts seen as using light touch for schools required to restructure. *Education Week*, pp. 1, 19.

This Ed Week article summarizes a conference held in Washington, DC to discuss how states and districts are responding to failing schools. Few students are taking advantage of the choice or services provided, and sanctions are not being strictly enforced. The papers discuss whether "gentler" options are working.

Palmer, S. & Coleman, A. (2004). *Blueprint for building a single statewide accountability system*. Washington, DC: Council of Chief State School Officers.

This publication includes a blueprint for developing a statewide accountability system within the confines of NCLB. It includes templates for determining how to meet adequate yearly progress, indicators, decision rules, and rewards and consequences. It provides a framework for considering both state and federal legal and policy contexts and lists the steps states should take in developing an accountability system that meets federal requirements.

Peterson, P. E., and West, M. R. eds. (2003). *No Child Left Behind? The Politics and Practices of School Accountability*. Washington, DC: Brookings Institution Press.

This edited book explores the research on school accountability and uses those results to predict the effects of No Child Left Behind. Contributors examine the law's origins, the political and social forces that gave it shape, the potential issues that will surface with its implementation, and the law's likely consequences for American education. It concludes that states' requirements for schools are likely to be softened over time by political opposition but that the end result may be enough to boost student achievement.

Peterson, P. E., and West, M. R. (2006). Is your school effective? *Education Next* (Vol. 2006 No. 4).

This article analyzed longitudinal data from Florida to compare average student growth in schools making AYP compared to growth observed in schools not making AYP. The authors found a small difference indicating the status measure was distinguishing among schools, but the size of the differences caused the authors to question the importance of such a distinction. Overall, the conclusion is that we need better measures to determine which schools are truly failing.

Porter, Andrew C. (1991). Creating a System of School Process Indicators. *Educational Evaluation and Policy Analysis, 13*, pp. 13-29.

This article offers a model of education inputs, processes, and outputs that includes a system of indicators and an analytic framework. An indicator system of school processes facilitates description, tracking reforms, and diagnosing school outputs.

Raudenbush, R. (2004). *Schooling, statistics, and poverty: Can we measure school improvement?* ETS Angoff Lecture, 2004.

This paper from an ETS lecture series compares two different approaches to measuring school quality and school improvement: status proficiency and value added. The

results draw attention to validity concerns with status measures but shows that both measures produce estimates with uncertainty and bias.

Rustique-Forrester, E. (2005). Accountability and the pressures to exclude: A cautionary tale from England. *Education Policy Analysis Archives, 13*(26). Retrieved February 13, 2007 from http://epaa.asu.edu/epaa/v13n26/.

This paper investigates the question of whether test-based rewards and sanctions produce the desired effects by examining the impact of accountability systems in England.

Stecher, B. & Hanser, L. (1992). Local accountability in vocational education: A theoretical model and its limitations in practice. MDS-291. Santa Monica, CA: RAND.

This report focuses on accountability in vocational education systems, but the principles are universal. They argue that an accountability system should contain four major components—goals, measures, a feedback loop, and a systemic change mechanism—and describe this model in detail.

Stecher. B. & Kirby, S. (2004) *Improving Educational Outcomes through Accountability: Lessons for Education from Other Sectors*. Santa Monica, CA: RAND.

This research brief analyzes accountability approaches from a variety of professions, including manufacturing, law, and health care and evaluates their usefulness for education.

Stecher, B. & Naftel, S. (2006). *Implementing state-based accountability (ISBA): Study design, state context, and accountability policies*. WR-380-EDU. Santa Monica, CA: RAND.

This paper, part of a symposium, summarizes a project designed to identify factors that enhance the implementation of standards-based accountability systems, foster change in school and classroom practice, and promote improved student achievement. The study followed three states that had various types of accountability systems in place prior to NCLB.

Stullich, S., Eisner, E., McCrary, J., & Roney, C., (Feb. 2006). *National Assessment of Title I, Interim Report. Volume I: Implementation*. Washington, DC: Policy and Program Studies Service, Office of Planning, Evaluation, and Policy Development.

This federally-funded evaluation of Title I programs contains key findings on the implementation of Title I under No Child Left Behind. It includes results of evaluation studies to detect the impact of educational practices and programs on student achievement.

Taylor, R., Beaudoin, J. & Goldschmidt, P. (2007). *Quality assurance practices associated with producing cohort graduation rates*. Washington, DC: Council of Chief State School Officers.

This report was written as a part of ASR SCASS work on quality control and validity of accountability indicators. Specifically linking data quality issues to accountability results from a practitioner's viewpoint, the paper lays out several quality assurance

steps states could implement at various points. Focusing specifically on the graduation rate indicator, this paper explores how graduation rates are calculated in many states, the need for calculation of a cohort graduation rate, and details several quality assurance practices states can use to produce valid and reliable cohort graduation rates.

Wilson, M. (Ed.) (2004). *Towards coherence between classroom assessment and accountability*. 103rd yearbook of the National Society for the Study of Education. Chicago, IL: University of Chicago Press.

This 103rd yearbook of the National Society for the Study of Education discusses the role of classroom assessment in accountability system. It includes a series of chapters by experts in the field of classroom assessment followed by a series of commentary chapters responding to those views. The topics focus on the challenges of designing and using tests to support the curriculum, inform the accountability system, and provide a framework to help teachers improve instruction. The link between statewide and classroom assessment is discussed.

Zimmer, R., Gill, B., Razquin, P., Booker, K., & Lockwood, J.R. (2007). *State and local implementation of the No Child Left Behind Act: Volume I—Title I school choice, supplemental education services, and student achievement*. Washington, DC: U. S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.

This federally-funded evaluation examines various sanctions applied to schools not making adequate yearly progress. Students from nine urban districts were followed to see how school choice or supplemental educational services affected student achievement. The study used a technique called a "student fixed-effects model" to compare the growth students made under the program to the progress they made in previous years in their regular public schools.

Appendix B: Matrix Synthesizing The Literature Review

Core Concepts	Guiding Questions	Key Elements	Options/Examples
Goals	What are the purposes of the accountability system? More specifically, what are the primary goals you are trying to accomplish with an accountability system? What accountability decisions will be made, and with what consequences?	Develop an explicit theory of action	Improve student learning by using assessment as a (a) tool for continued improvement or (b) as a means for delivering rewards or sanctions
	What are the main contexts, political or otherwise? What are the main legal and policy constraints or specifications?	Understand the legislative, policy, and legal influences and understand how those considerations will determine the design and implementation of the accountability system.	
	Are the long-term and short-term goals of the system worthwhile, realistic, and achievable?		
Performance Indicators	What data are available and will be used in the accountability system?	Accountability systems should employ different types of data from multiple sources.	Consider multiple measures of achievement such as incorporating both teacher-developed assessments and standardized assessments into accountability systems.
			In addition to test scores, include teacher qualifications, curriculum, and learning environments.
		Decisions about individual students should not be made on the basis of a single test.	Tests should be used as one component of an accountability system, not as the system itself
		Accountability systems should include data elements that allow for interpretations of student, institution, and administrative performance.	Eliminate artificial barriers to students demonstrating ability, such as time limits and inflexible administration
	Which students should be included in the system and when should they be tested?	Accountability systems should include the performance of all students, including subgroups that historically have been difficult to assess.	Consider whether to use all grades or selected grades. Special populations? New entrants?
			Decide whether to require each subgroup to meet a separate bar or the same bar
	How do you ensure the data for the indicators are interpreted correctly and/or produce valid results?	The validity of the measures that have been administered as part of an accountability system should be documented for the various purposes of the system.	
		If tests are claimed to measure content and performance standards, analyses should document the relationship between the items and specific standards or sets of standards.	Alignment studies
	What technical issues and additional analyses will need to be addressed in developing a valid set of indicators?	Multiple test forms should be used when there are repeated administrations of an assessment.	
		If tests are to help improve system performance, there should be information provided to document that test results are modifiable by quality instruction and student effort.	Performance components should provide "tests worth teaching to"

Core Concepts	Guiding Questions	Key Elements	Options/Examples
Design Decisions	What are the schools/students (or others) to be held accountable for?	Match the indicators and decision rules to the goals	Process indicators, performance indicators, monetary indicators
	Who is held responsible?	Develop a system of shared responsibility.	Information must be available both on students (achievement) and adults (instruction)
	What accountability model best serves your purpose? (status, improvement, or growth)	Deliberately consider status versus improvement versus growth and align the model with the goals	Ensure accountability system can distinguish among schools that are improving and those that are not.
	How do we move from an old accountability system to a new one?	Consider previous policies when making new. Should new replace or supplement old?	E.g., moving from process-based accountability to outcome-based
	How will data from multiple measures and indicators be combined to categorize schools and make an accountability judgment?	The weighting of elements in the system, including different types of test content, and different information sources, should be made explicit.	Decide whether your system will use a composite score or conjunctive rules
			Consider the use of an index to weight elements appropriately. For example an index may include achievement test results (normed and standardized tests), participation rates, attendance rates, and graduation rates, each weighted according to the values of the system.
			Complexity and trade-offs with fairness (complex formulas to create indices can include all relevant factors but be difficult to communicate to the public)
	What is satisfactory progress?	Proceed at a pace that is technically defensible and politically reasonable	Schools should be accountable for the value they add to student learning, not the effects of prior instruction; school systems should be accountable for the cumulative learning of students over their career in the system.
			Decide between letting the starting point for a school determine the goal or setting the same goal for all schools
			Determine the end point as a fixed point in time, or let the annual goals determine the end point for the final goal

Core Concepts	Guiding Questions	Key Elements	Options/Examples
Design Decisions (continued)	How do we balance reliability and validity concerns in making decisions about schools?	Make sure the overall design of the system is valid and that it is implemented in a way that leads t valid results. Consider all threats to validity and eliminate those as a first step.	Carefully consider the real costs of Type I and Type II errors and attempt to reach a reasonable balance between the two. Do not assume that the costs associated with either type of error are insignificant.
		Use valid assessments, even if higher assessment validity comes at the cost of lower student-level reliability. Lower student-level reliability will have only a minor impact on the reliability of school scores.	
		Do not set a fixed N for subgroups. There is no choice of fixed N that will provide for both valid and reliable results, and most choices provide for neither.	
		Do not attempt to measure school improvement over 1 year. The uncertainty is too great relative to the amount of improvement expected to make reliable judgments for most schools.	
		Rules for determining adequate progress of schools and individuals should be developed to avoid erroneous judgments attributable to fluctuations of the student population or errors in measurement.	
System Monitoring and Evaluation	How will the accountability system be monitored and evaluated?	Longitudinal studies should be planned, implemented, and reported evaluating effects of the accountability program.	
	How can the intended and unintended effects of the accountability system be evaluated? (monitoring system with ongoing evaluation)	The validity of test-based inferences should be subject to ongoing evaluation.	
	How will the system design incorporate the need for revisions over time?	Develop steps to continuously evaluate and improve the effectiveness of the accountability system itself.	
	To what degree does the system support high-quality instruction and student access to education and minimize unanticipated outcomes?	School districts must collect and analyze information about the effectiveness of educational programs	
	How will we determine if the consequences are effective?		Monitor the effectiveness of incentives for teachers.
Communication	How will accountability results be reported?	Reports to districts and schools should promote appropriate interpretations and use of results.	Provide data to schools in a way that they can learn to use it.
		Results must be reported in writing with time for appeal.	Consider context-sensitive reporting that promote awareness of the invludences affecting outcomes.
		Reports need to clearly communicate results and include all relevant data.	Include multiple indicators of performance, error estimates, and performance by subgroups

Core Concepts	Guiding Questions	Key Elements	Options/Examples
Communication (continued)	How will parents and the general public be informed as to the goals and limitations of the system?	Accountability expectations should be made public and understandable for all participants in the system.	
		System results should be made broadly available to the press, with sufficient time for reasonable analysis and with clear explanations of legitimate and potential illegitimate interpretations of results.	
Support	How will the state provide the schools with the support needed to improve the results required in the accountability model?	Reciprocity of accountability and capacity	Pilot school program; teacher training; local autonomy; financial support; consulting support
		Develop an infrastructure to provide expertise and support at the local levels	Support schools through TA, financial support
			Access to curriculum-related materials; professional development planning process; systematic school improvement process
		Create incentives to stabilize the population of teachers and administrators in low-performing schools	Financial (loan forgiveness; extra compensation); Professional development (access to courses and advanced degree programs); Improved working conditions (reduced teaching loads; improved physical infrastructure; better materials)
		Require school districts to develop systemwide improvement strategies	1. Needs assessment and planning (revisions of district plans, data analysis tools to identify where to target resources) 2. Data analysis (state provides the analysis or tools to help districts manage large data sets) 3. Capacity building (focusing on enhancing districts' PD, leadership, or curricular activities) 4. Resource allocation (aligning budgets with priorities, providing tools to assist districts in managing and targeting funds) 5. Progress monitoring (conducting district audits and onsite reviews)
	What are the actual costs incurred by the system and what are the necessary tradeoffs between quality and cost?	Allocate resources where they are needed most.	Consider state capacity for remedy (don't underestimate costs for sanctions)
	What support (e.g., professional development) do teacher and administrators need to implement the system?	Develop plan for supports	Communicate clear, high expectations to students; align PD to district's mission; provide teachers sufficient time for planning; provide students with OTL; implement appropriate instructional interventions
			Require and fund diagnostics for students who are not succeeding
	What will you do about the problems uncovered through the accountability system?	Policy retreats reduce credibility and public support	School board, superintendent, & treasurer create conditions for a school district's success

Core Concepts	Guiding Questions	Key Elements	Options/Examples
Consequences	What consequences can your accountability system support? What rewards & sanctions are effective in obtaining the goals of your accountability system?	Stakes for accountability systems should apply to adults and students and should be coordinated to support system goals.	
		Accountability systems should begin with broad, diffuse stakes and move to specific consequences for individuals and institutions as the system aligns.	
		Individual and collective stakes should be based on defensible, empirically-based theories about what is possible to accomplish on measured performance within a given period of time.	
	When and how will the consequences be applied?	Appeal procedures should be available to contest rewards and sanctions.	Opportunity to appeal decisions enhances the fairness and transparency of accountability systems.
		Decision rules and order and timing of consequences should be clear.	
	Who will the consequences affect?	Consider student incentives (is it fair to penalize teachers for results when students are not held accountable? Motivation issue for students.)	Separate consequences for students, schools, and districts
	How will you ensure the consequences are applied appropriately and effectively?	Stakes for results and their phase-in schedule should be made explicit at the outset of the implementation of the system.	
		Stability and credibility (educators must believe rewards and sanctions will happen)	

	AEA, 2006		Goertz, 2001
	AIR, 2006		Gong & ASR SCASS, 2002
	Ananda & Rabinowitz, 2001		Hamilton & Stecher, 2004
	Baker, Linn, Herman, & Koretz, 2002		Hill & DePascale, 2003
	Darling-Hammond, 2006		Linn, 2003
	Elmore, 2003		Linn, 2005b
	Elmore, 2004		O'Day, 2002
	Erpenbach, 2002		MA Diagnostic Fact Finding Review
	Forte Fast & Hebbler, 2004		ODE Strategic Plan
	Fuhrman, 1999		Combined or edited indicators

Appendix C: Checklist

The following checklist was created to help state and district-level policymakers evaluate the design of their accountability system. A checklist approach seemed most user-friendly as it highlights the considerations states should undertaking during either the development of a new accountability system or an evaluation of a current one.

Checklist for State Accountability Systems

Goals

- *What are the purposes of the accountability system? Why are you implementing the system?*
  - ☐ Explicate the local, state, and federal requirements that this system is trying to fulfill.
- *What are the primary goals you are trying to accomplish with an accountability system? What are you hoping to accomplish?*
  - ☐ Establish and prioritize goals.
- *What accountability decisions will be made and with what consequences?*
  - ☐ Develop an explicit theory of action that relates the goals and design of the system to anticipated outcomes by the state, schools, and others.
- *In what systems are you working? What are the main legal and policy constraints or specifications?*
  - ☐ Explore the legislative, policy, and legal influences and understand how those considerations will determine the design and implementation of the accountability system.
- *In broad terms, what are the schools/students (or others) to be held accountable for?*
  - ☐ Match the indicators and decision rules to the goals.
  - ☐ Begin to consider how to combine multiple indicators to match the goals by determining whether the goals best fit a status, improvement, or growth model.

Performance Indicators

- *What data are available and will be used in the accountability system?*
  - ☐ Employ different types of data from multiple sources.
  - ☐ Think broadly about types of measures to include, such as measures of teacher qualifications, curriculum, and learning environments.
  - ☐ Include data elements that allow for interpretations of student, institution, and administrative performance as appropriate.
  - ☐ Avoid making decisions about individual students, educators, or schools on the basis of a single indicator.
- *Which students should be included in the system?*
  - ☐ Include the performance of all students, including subgroups that historically have been difficult to assess.
  - ☐ Determine how to include special populations.
  - ☐ Consider whether to use all grades or selected grades.

- ☐ Determine how to include alternative education programs (such as vocational schools or schools intended for short-term attendance by students with disciplinary problems).
- ☐ Determine whether to include students new to the school or district.

- *When should indicators be collected?*
  - ☐ Consider when you need the data to make accountability decisions.
  - ☐ Decide whether your system's goals support fall testing, spring testing, or more frequent testing.
  - ☐ Eliminate artificial barriers to students demonstrating ability, such as time limits and inflexible administration.

- *How do you ensure the data for the indicators are interpreted correctly and/or produce valid results?*
  - ☐ The accountability system is only as strong as its indicators.
  - ☐ The validity of the measures that have been administered as part of an accountability system should be documented for the various purposes of the system.
  - ☐ Develop a quality assurance plan that examines the accuracy and validity of all indicators.

- *What technical issues and additional analyses will need to be addressed in developing a valid set of indicators?*
  - ☐ If tests are to help improve system performance, document that test results are modifiable by quality instruction and student effort.
  - ☐ Performance components should provide appropriate representation of the knowledge and skills valued by the state and/or district.
  - ☐ Reliability of other indicators used as outcome measures, such as teacher qualifications and graduation rates, should be documented.

Design Decisions

- *Who is held responsible?*
  - ☐ Develop a system of shared responsibility.
  - ☐ Consider providing information both on students (achievement) and educators (instruction).

- *What accountability model best serves your purpose?*
  - ☐ Deliberately consider status versus improvement versus growth and align the model with the goals.
  - ☐ If the goals imply treating low performing but improving schools differently from low performing schools, ensure accountability system can distinguish among schools that are improving and those that are not.

- *How will data from multiple measures and indicators be combined to categorize schools and make an accountability judgment?*

- □ Make explicit the weighting of elements in the system, including different types of test content and different information sources.
- □ Decide whether your system will use compensatory or conjunctive rules (both at the individual level and in the aggregate) by considering the goals and theory of action.
- □ Consider the use of an index to weight elements appropriately.
- □ Consider the complexity and trade-offs with fairness (complex formulas to create indices can include all relevant factors but be difficult to communicate to the public).

- *What is satisfactory progress?*
  - □ Proceed at a pace that is technically defensible and politically reasonable.
  - □ Decide whether to hold schools accountable for how they increase student learning or to the absolute level of learning (which includes effects of prior instruction).
  - □ Consider holding school systems (e.g., districts) accountable for the cumulative learning of students over their career in the system.
  - □ Decide between having the starting point for a school determine the goal or setting the same goal for all schools.
  - □ Decide between determining the end point as a fixed point in time, or allowing the annual goals determine the end point for the final goal.
  - □ Decide whether each subgroup should meet a separate bar based on their starting point or the same bar.

- *How do we balance reliability and validity concerns in making decisions about schools?*
  - □ Strive to make the overall design of the system valid and implement it in a way that leads to valid results. Consider all potential threats to validity and eliminate those as a first step.

  - □ Carefully consider the real costs of Type I and Type II errors and attempt to reach a reasonable balance between the two. Do not assume that the costs associated with either type of error are insignificant.
  - □ Consider the goal of inclusion of all subgroups when making determinations of a minimum N for subgroups.
  - □ Explore the uncertainty involved in measuring school improvement over one year in terms of making reliable judgments for most schools.
  - □ Develop rules for determining satisfactory progress of schools and individuals to avoid erroneous judgments attributable to fluctuations of the student population or errors in measurement.

- *How do we move from an old accountability system to a new one?*
  - □ Consider previous policies when making new. Decide whether the new system should replace or supplement the old by considering all implications.
  - □ Determine how to transition any school categorizations from the old system to the new.

- □ Include a reasonable timeline for the transition.

Consequences

- *What consequences (positive and negative) can your accountability system support?*
  - □ Determine the appropriate consequences for your goals (remember that providing appropriate support in response to an identified need is a consequence).
  - □ Begin with broad, diffuse stakes and move to specific consequences for individuals and institutions as the system aligns.
- *What rewards & sanctions are effective in obtaining the goals of your accountability system?*
  - □ Base individual and collective stakes on defensible, empirically-based theories about what is possible to accomplish on measured performance within a given period of time.
  - □ Coordinate stakes for accountability systems to support system goals.
- *Who will the consequences affect?*
  - □ Consider student incentives (is it fair to have consequences that affect teachers if students are not motivated to do their best?).
  - □ Consider separate consequences for students, educators, schools, and districts.
- *When and how will the consequences be applied?*
  - □ Clearly state your decision rules and the order and timing of consequences.
  - □ Make appeal procedures available to ensure valid and reliable decisions.
- *How will you ensure the consequences are applied appropriately and effectively?*
  - □ Make explicit the stakes for results and their phase-in schedule at the outset of the implementation of the system.
  - □ Create stability and credibility (educators must believe rewards and sanctions will happen).
- *How will you determine if the consequences are effective?*
  - □ Monitor the effectiveness of any rewards (e.g., incentives for teachers) or sanctions (e.g., restructuring schools) on performance.

Communication

- *How will accountability results be reported?*
  - □ Ensure the reports to districts and schools promote appropriate interpretations and use of results.
  - □ Provide data to schools in a way that they can learn to use it.
  - □ Monitor the latest research in effective reporting and continually review the reporting system.
  - □ Ensure reports clearly communicate results and include all relevant data.

- □ Include multiple indicators of performance broken out by subgroups in the reports.
- □ Make error estimates of all measures available.
- □ Ensure all reports meet FERPA regulations (e.g., do not report individually identifiable data)
- □ If there are consequences associated with the results, the results must be provided to local school systems with time for appeal before results are released to the general public.

- *How will parents and the general public be informed as to the goals and limitations of the system?*
  - □ Make accountability expectations public and understandable for all participants in the system.
  - □ Prepare information guides specific to the audience with various levels of detail as appropriate.
  - □ Make system results broadly available to the press, with sufficient time for reasonable analysis and with clear explanations of legitimate and potential illegitimate interpretations of results.

Support

- *How will the state provide the schools with the support needed to improve the results required in the accountability model?*
  - □ Acknowledge the reciprocity of accountability and capacity.
  - □ Develop an infrastructure to provide expertise and support at local levels.
  - □ Create incentives to stabilize the population of teachers and administrators in low-performing schools.
  - □ Encourage school districts to develop systemwide improvement strategies.

- *What are the necessary tradeoffs between quality and cost?*
  - □ Determine the actual costs incurred by the system.
  - □ Allocate resources where they are needed most.
  - □ Consider state capacity for remedy.

- *What support do administrators need to implement the accountability system?*
  - □ Develop plan for supports.
  - □ Communicate clear, high expectations to students.
  - □ Provide resources for professional development aligned to district's mission.
  - □ Consider diagnostics for students who are not succeeding.

- *What support (e.g., professional development) do teachers need to implement the system?*
  - □ Align professional development to district's mission.
  - □ Provide teachers with sufficient time for planning.
  - □ Provide students with the opportunity to learn.
  - □ Implement appropriate instructional interventions.

- *To what degree does the system support high-quality instruction and student access to education?*
    - □ Collaborate with appropriate stakeholders to create conditions for a school or school district's success.
    - □ Consider involving schools and school districts in collecting and analyzing information about the effectiveness of educational programs.
    - □ Document and disseminate success stories where high-quality instruction and student access to education were maximized.

System Monitoring, Evaluation, and Improvement

- *How will the accountability system be monitored and evaluated?*
    - □ Plan for annual monitoring and evaluation of the system.
    - □ Consider state capacity for monitoring.

- *How can the intended and unintended effects of the accountability system be evaluated?*
    - □ Plan and implement longitudinal studies evaluating the effects of the accountability program.
    - □ Determine whether unintended effects of the accountability system are positive or negative and minimize unintended negative effects.
    - □ Conduct ongoing evaluations of the validity of the indicators.
    - □ Consider using an external evaluator to independently review the system and provide confirming (or disconfirming) evidence to supplement the internal evaluation

- *How will the results of this monitoring and evaluation be used to improve the system?*
    - □ Avoid frequent or drastic changes to the accountability system as they reduce credibility and public support.

- *How will the system design incorporate the need for revisions over time?*
    - □ Develop steps to continuously evaluate and improve the effectiveness of the accountability system itself.
    - □ Take steps to evaluate and improve the theory of action as well as any of the components.
    - □ Ensure the system is sufficiently flexible to support additional indicators as new data become available.

    - □ How will the system design incorporate the need for revisions over time?
    - □ Develop steps to continuously evaluate and improve the effectiveness of the accountability system itself.
    - □ Take steps to evaluate and improve the theory of action as well as any of the components.
    - □ Ensure the system is sufficiently flexible to support additional indicators as new data become available.

Appendix D: Specific State Examples

Maine Rubric for School Improvement

The second example shows the rubric developed by the Maine Department of Education that schools can use to help determine their capacity to do school improvement work.

Maine School Improvement Rubric September 2004

I. Commitment, Planning, Organization

	Ineffective (1)*	Effective (3)*	Most Effective (5)*	Comments
Awareness	Few staff members display knowledge and understanding of the school improvement process (SIP)	Many staff members know about the purpose and procedures of the SIP.	All or nearly all staff members know and understand the purpose and procedures of the SIP..	
Commitment	Few staff members support the school improvement process.	Many staff members support the school improvement process.	All or nearly all staff support the school improvement process.	
Roles and Responsibilities	Few staff understand or support the leadership structure of the school improvement process.	Many staff understand and support the leadership structure of the school improvement process.	All or nearly all staff understand and support the leadership structure of the school improvement process.	
Committee Composition	Membership of steering committee lacks rationale for selection, is not representative of all groups, or consists of volunteers only.	Steering committee membership is representative of staff and administration.	All groups are represented in the committee structure, including (when appropriate) community and student body.	
Timeline	The timeline identifies only the end of the 5-year cycle and the external team visit.	The timeline also identifies dates for significant events and activities of the SIP.	The timeline identifies specific dates for completion of each phase or activity.	

	Ineffective (1)*	Effective (3)*	Most Effective (5)*	Comments
Mission Statement Development	Few staff members and no community members participate in the development of the mission statement.	Many staff members and some community members participate in the development of the mission statement.	All staff members and many community members participate in the development of the mission statement.	
Mission Statement	The mission statement is lengthy, jargon-filled, and apparently unconnected to decision-making.	The mission statement is clear and direct, but its connection to decision making is not always apparent	The mission statement is memorable and provides a rationale for many of the decisions in the improvement process	
Leadership	Steering committee meets irregularly and exerts little or no control over process	Steering committee meets on a regular schedule and reports progress to administration.	Steering committee meets on a regular schedule, coordinates and monitors improvement activities, and keeps all staff members informed.	

II. Gathering And Analyzing Data

	Ineffective (1)*	Effective (3)*	Most Effective (5)*	Comments
Community Data	The profile committee collects no community information or collects irrelevant and unorganized community information and does not relate it to school improvement	The profile committee collects, organizes, and presents community information relevant to the SIP.	The profile committee collects, organizes and clearly presents community information, with an analysis of the relation of that information to the SIP.	
Student Demographic Data	The school profile contains little or no student demographic data.	The school profile contains longitudinal demographic data, including projections for future enrollments, and an analysis of the trends.	The school profile contains comprehensive, disaggregated, and longitudinal demographic data including subject enrollment trends and an analysis of the data.	
Student Performance Data	The collection of student performance data meets state and federal requirements but lacks analysis (or bases analysis based on minimal data)	Multiple sources of longitudinal performance data (NRT, CRT, classroom assessments, etc.) are clearly organized, disaggregated where appropriate, graphically displayed and at least superficially analyzed.	In addition to a comprehensive collection and clear presentation of performance data collected over time, the school profile contains a thoughtful analysis of the implications drawn from that data.	

	Ineffective (1)*	Effective (3)*	Most Effective (5)*	Comments
Data Collection and Reporting Process	The school profile contains limited or irrelevant data collected irregularly from a few sources and lacks organization.	Data collection is an on-going process drawing from many sources. Information is clearly organized and graphically displayed.	The data collection process is comprehensive, continuous, and focused on trend data. Results are appropriately disaggregated, clearly presented, and thoroughly analyzed.	
Data Display	Display of data is unclear, unexplained, and difficult to interpret.	Data is displayed in clearly labeled graphs and charts.	Data is displayed in clearly labeled graphs and charts accompanied by an explanatory narrative.	
Use of Data for Decision-making	School profile data appears unrelated to school improvement decisions.	School profile data is analyzed and is generally used to guide goal selection.	The school prepares a description of its strengths and needs based on the school profile. A process of applying these conclusions to decision-making and goal setting is documented. Further diagnostic data needs (if any) are identified.	

Council of Chief State School Officers
Washington, DC