UNDERSTANDING HOW PRINCIPALS USE DATA IN A NEW ENVIRONMENT OF ACCOUNTABILITY

Prepared by
Kerry Englert
Dawn Fries
Bryan Goodwin
Mya Martin-Glenn
Sheri Michael

October 30, 2004

McREL

TABLE OF CONTENTS

INTRODUCTION

This report summarizes the findings of the second in a series of three studies that examine how educators at various levels use assessment data and accountability systems. The information presented in these studies focuses on understanding how educators incorporate effective assessment and accountability practices into their work. The first study examined practices that superintendents use in their districts to encourage and foster a climate of using data from assessment and accountability systems to inform decision making related to allocating resources, selecting curricula, evaluating personnel, and improving the education systems (see Englert, Fries, Goodwin, & Martin-Glenn, 2003). The present study expands that work by surveying principals in the same districts in order to understand how assessment and accountability practices are actualized in schools. The final study in this series will examine teacher responses to these policies and practices and link the three studies together to examine the consistency and cohesiveness across the three levels of educators.

The purpose of this study is to develop a better understanding of the assessment and accountability practices and policies principals are implementing at the school level and to determine whether those policies are associated with school and student improvement. This study generated descriptive information about the needs schools and districts have for effectively using data, how schools and districts are using data to guide classroom practice, and whether there are differences in data use based on the level of school proficiency.

This report serves as the overall technical report on the study findings and can be used to develop an understanding of principals' perceptions about the use of data and accountability information in schools. Future brief reports will highlight specific issues of interest to a variety of audiences. One report might be directed to personnel at state departments of education to understand how data are being used in districts and schools and what practices tend to be associated with schools that are improving and/or high achieving. Another report might be directed to school and district administrators to help them examine their current practices and how these practices relate to elements of good assessment and accountability practices.

SEVEN CHARACTERISTICS OF EFFECTIVE ACCOUNTABILITY SYSTEMS

The effectiveness of schools' accountability policies is of paramount concern to the success of No Child Left Behind Act (NCLB). If schools are not actively engaged in effectively using accountability data, generating the increases in student achievement required by this legislation becomes unattainable. The current attention given to accountability in education has resulted in an increased focus on assessing and monitoring student achievement. Many education researchers and scholars cite both positive and negative impacts of this change in focus. On the positive side, increased accountability is credited with focusing schools on learning outcomes, emphasizing efforts that help all students learn, providing staff with information to adjust curricula or staffing, and informing parents about how well their children and their schools are performing. Others cite negative impacts of increased accountability such as a narrowing of the curriculum based solely on test content, increased retention rates, increased numbers of students placed in special education programs, and an increased focus on test performance rather than on other important factors such as school safety, dropout rates, and discipline (see Goodwin, Englert, & Cicchinelli, 2003).

Goodwin et al. (2003) reviewed the literature on accountability systems and identified 12 frequently cited characteristics of good systems that are essential to examine when evaluating accountability systems[1]. Seven of these were determined to be most relevant to principals (see Englert et al., 2003): (1) high expectations for all students; (2) high-quality assessments aligned with standards; (3) alignment of resources, support, and assistance for improvement; (4) sanctions and rewards linked to results; (5) multiple measures; (6) diagnostic uses for data; and (7) readily understandable to the public. These elements were selected and used as the foundation for constructing a survey of principals. Grounding the research study in these characteristics allows for an examination of the overall effectiveness of current school policies and practices. The selected characteristics are discussed in more detail in the following section.

High Expectations for All Students

An important purpose of accountability systems is to ensure all students have equal access to opportunities for learning. One way to promote this equity is to set high expectations for all students (Goodwin et al., 2003). Grissmer and Flanagan (1998) found particularly salient examples of high expectations in both the Texas and North Carolina state accountability systems. The accountability policies in both states emphasized the expectation that *all* students should meet standards. These policies called for teachers and administrators to hold students, both advantaged and disadvantaged, to the same high standards.

One way to demonstrate that all school staff members have the same high expectations for all students is to test all students. By testing all students, teachers can focus their efforts on making sure that each student is learning at the desired level. According to Baker, Linn, Herman, and Koretz (2002), "accountability systems should include the performance of all students, including subgroups that historically have been difficult to assess" (p. 2). Research has found that maintaining and communicating a school-wide academic press for achievement is a contributing factor to helping low-achieving students achieve higher standards and consequently creating an effective school (Teddlie & Reynolds, 2000; Creemers, 1994; Marzano, 2000). This notion of high expectations is a cornerstone of the NCLB legislation because NCLB directly focuses on assessing all students in order to draw attention to subgroups of students who might not be meeting performance standards.

High-quality Assessments Aligned with Standards

Although many districts use multiple data sources in order to more accurately account for and gauge their progress, student achievement, as measured through statewide assessments, is becoming synonymous with accountability and assessment data is the primary source of information under NCLB. Assessments have become the primary tool for gauging students' success as well as the success of teachers, schools, and districts. Because of the central role that assessment instruments play in accountability systems, it is necessary to ensure that they are of high quality. After specifying clear

---

[1] The 12 essential characteristics identified by Goodwin et al. (2003) were (1) clear standards and expectations; (2) high expectations for all students; (3) high-quality assessments aligned with standards; (4) alignment of resources, support and assistance for improvement; (5) sanctions and rewards linked to results; (6) multiple measures; (7) diagnostic uses for data; (8) readily understandable to the public; (9) flexibility and fairness to allow for local differences and creativity; (10) balanced, comprehensive design; (11) stakeholder support/engagement; and (12) fairness provisions.

academic standards, authors of most accountability systems contend that student progress on standards needs to be measured using assessments that are aligned to those standards (Guth et al., 1999; Baker et al., 2002; Buckendahl, Impara, & Plake, 2002). Moreover, most accountability standards and models also heavily emphasize the quality of these examinations. Thus, the most relevant aspects of quality for the surveyed population were deemed to be alignment of the assessment to state standards, the technical quality of the assessment instrument, and the provision of useful data to stakeholders.

Alignment of Resources, Support, and Assistance for Improvement

Many models of accountability emphasize the need to align resources and support with the goals of the system. Grissmer and Flanagan (1998) note that "recent research is now supporting the hypothesis that resource levels can make significant differences in achievement, and that disadvantaged students probably benefit more from increased educational resources" (p. 24). As a result, "the distribution of resources among schools and school districts needs to be perceived as fair and equitable for an assessment and accountability system to function effectively" (p. 24).

Researchers have argued that in addition to ensuring equitable funding, states must help schools develop the capacity to make necessary improvements and the flexibility to direct the funds toward specific programs that target areas in need of improvement. New York City schools, for example, which were given the authority to align resources with their instructional improvement plans, saw a small yet significant increase in student test scores (Siegel & Fruchter, 2002). Grissmer, Flanagan, Kawata, and Williamson (2000) found that states with a lower average socioeconomic status (SES) can show significant gains in student achievement "through modest increases in resources, if allocated to specific programs" (p. 101). Researchers at Wested (Guth et al., 1999) made the same assertion, noting that schools identified as not meeting performance goals should have some avenue of assistance to help them improve. Fuhrman (1999) pointed out that although research "shows that new accountability systems can be motivating . . . .the mere imposition of a new accountability system . . . does not unleash some hidden capacity" (pp. 8–9). In short, though accountability systems may create more *will*, they often fail to provide schools with the *way* (i.e., tools and strategies) to succeed. In light of such concerns, Sirotnik and Kimball (1999) argue that

> the accountability system must include support for and monitoring of substantial, long-term professional development opportunities for teachers and administrators to inquire into their disciplines and to review and revise their pedagogical content knowledge and teaching and leadership skills (including evaluation and assessment). (p. 211)

Proper training and skills in assessment practices are vital because teachers can spend from a third to a half of their time engaged in assessment-related activities (Stiggins, 1988, as cited in Trevisan, 1999). Therefore, district administrators have an important responsibility to teachers as well as school administrators to provide leadership on assessment issues. A study by Trevisan (1999) found that only "14 states have assessment requirements for superintendents" (p. 9). Even with these requirements in place, there are no guidelines in these states for specifying or developing the needed competencies.

Sanctions and Rewards Linked to Results

There are many examples of accountability models that use sanctions and rewards to spur school improvement efforts, despite the lack of clear evidence that either strategy is effective (see Goodwin et al., 2003). Most notably, NCLB mandates increases in schools' performance over time. In a report examining the gains in elementary and junior high school test scores in Texas and North Carolina, Grissmer and Flanagan (1998) found that both states "have financial rewards for schools based on performance, and have the power to disenfranchise school districts and remove principals based on

sustained levels of poor performance" (p. iv). Walberg (2002) argued for the use of sanctions and rewards such as those evident in Texas and North Carolina — sanctions and rewards that move beyond publicly highlighting school successes and failures — because "simply publishing results appears insufficient for progress. … Schools of choice risk closing if they attract no students. Analogous thinking dominates much of society. Why not schools?" (p. 159). Accordingly, *Education Week*'s (2002) ratings of state accountability systems includes whether the state "holds schools accountable for performance" (p.12) through rewards and sanctions, including closure, reconstitution, transfers, and withholding of funds.

Sanctions and rewards are listed in the National Center for Research on Evaluation, Standards, and Student Testing's (CRESST) *Standards for Educational Accountability Systems* with the advisement that they should start out broad and diffuse, then "move to specific consequences for individuals and institutions as the system aligns" (Baker et al., 2002, p. 5). This allows stakeholders the opportunity to make the changes necessary to meet expectations. Baker et al. caution, however, that "if test data are used as a basis of rewards or sanctions, evidence of technical quality of the measures and error rates associated with misclassification of individuals or institutions should be published" (p. 3). They also stress the importance of having an appeal procedure in place so that individuals have a way to address possible extenuating circumstances.

However, not all educators support the notion of linking sanctions to accountability data. Sirotnik and Kimball (1999), for example, disagree with linking progress to sanctions and rewards in any fashion. In their view "the accountability system must not be punitive, either to students or to their teachers and schools" (p. 213). After reviewing principles of learning theory, Sirotnik and Kimball found that punishment did not seem to be effective in changing behavior. It would be more productive, they assert, to support districts and schools that are not making adequate progress than to punish them.

Multiple Measures

Many accountability guidelines call for the use of multiple measures in order to ensure a more complete and accurate assessment of students, teachers, and schools (Sirtonik & Kimball, 1999; Baker et al., 2002). Sirtonik and Kimball argue that an "accountability system must not be driven by a single indicator (e.g., test scores) and simplistic formulas for rewards or sanctions based on that indicator" (p. 211). Similarly, researchers at CRESST maintain that "decisions about individual students should not be made on the basis of a single test" (Baker et al., 2002, p. 3). They suggest that it "is important to consider other student outcome data such as attendance, mobility, and rates of retention in grade, dropout and graduation" (p. 2). Other measures that can be useful in evaluating schools include school safety data, parent feedback, expulsion rates, principal feedback, and student performance (as measured by course grades and homework).

District and School Personnel Data Usage[2]

Measuring student achievement and school progress toward goals in multiple ways is important; equally critical is strategically *using* these data to diagnose problems and work toward solutions. Research has found that in fact principals do want to use data to provide instructional leadership in a more informed way (Torrence, 2002). By using data to evaluate curricula, staff, and students, principals can focus their efforts and resources in the areas that are the most deficient. Conversely, Fuhrman (1999) found that

---

[2] In the Goodwin et al. (2003) report, this element is entitled "diagnostic uses for data."

many schools labeled as "inadequate" did not use test results diagnostically. The label created a state of urgency, but educators had little guidance about where to channel their urgency.

In order to provide schools and teachers with diagnostic information, the data must be specific enough to be useful in identifying and correcting problems. CRESST's *Standards for Educational Accountability Systems*, for example, calls for accountability systems to "include data elements that allow for interpretations of student, institution, and administrative performance" (Baker et al., 2002, p. 2). Including data from all levels facilitates the process of identifying areas of improvement in each area (American Education Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). The key, according to Reeves (2002), is providing school communities with sufficient data so that they can determine what works. Accordingly, it's important to consider different types of data from a range of sources at the state, district, and school levels.

Informative to Parents and the Community[3]

Another commonly identified element of accountability systems is their ability to inform students, parents, and teachers about student progress and status. Furthermore, providing information to parents and community members is a key requirement of NCLB. The legislation requires that parents be given information about the performance of their child's school promptly and that the information be clear and understandable. In order to do that, many argue, accountability systems should produce data and publish results that are meaningful to all interested parties. In its evaluation of state accountability systems, the Princeton Review (2002), for example, checks to see if the "performance data [are] shared with the public along with explanation and contextual detail appropriate for a general audience" (p. 6).

According to Walberg (2002), one principle that encourages the effective use of accountability system data is "user-friendliness" (p. 158). Reports about student and school progress should be useful and understandable to all interested parties regardless of their level of technical knowledge. It would be most effective to report the data in a variety of forms so that parents, school board members, and educators can readily make sense of the information. "What isn't as useful," argues Walberg, "is a mass of undigested numbers often reported by states and districts in large, unwieldy books of computer printouts" (p. 158).

Although it may be helpful to members of the public to combine data into a single, or small, set of numbers, CRESST researchers argue that when this approach is taken, states need to make explicit "the weighting of elements in the system, including different types of test content, and different information sources" (Baker et al., 2002, p. 2). They also recommend that "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" (p. 5).

The seven characteristics discussed here relate specifically to issues that affect principals or issues that principals can play a direct role in implementing or affecting. The No Child Left Behind Act has focused assessment and accountability issues in several areas such as ensuring that schools communicate with parents, and encouraging schools to have high expectations for all students. This focus is coupled with research that suggests that principals want to use data to direct teaching and provide instructional leadership (Torrence, 2002). The background literature helped focus the current study on the following research questions.

---

[3]In the Goodwin et al. (2003) report, this element was entitled "readily understandable to the public."

RESEARCH QUESTIONS

In order to understand principals' use of assessment and accountability systems, the following research questions guided this study:

- What policies and practices are principals using to meet new accountability demands?

- Are principals using policies and practices that research and literature have identified as being effective?

- Do policies and practices of high-performing schools differ from those of low-performing schools?

METHOD

In order to provide evidence answers to these three research questions, a survey was developed to measure the degree to which principals incorporated the above characteristics into their accountability policies and practices. The section that follows describes the development of the survey including the incorporation of the seven characteristics of effective accountability systems described in the previous section.

Instrument Development

The survey was developed based on the seven elements derived from the literature, and school and principal background information. Closed- and open-ended items were developed that were aligned with each of the seven characteristics. The survey measured principals' perceptions and policies on accountability through closed-ended questions and generated detailed contextual information about these elements via open-ended items. Using both of these formats resulted in a more complete picture of the policies and practices principals were advocating. Measures[4] were developed from the closed-ended items to provide more reliable indicators of the constructs and to ensure that principals' practices related to each of the elements were more accurately interpreted.

Prior to distributing the survey to principals, the survey was pilot-tested with a small group of four former principals and school administrators. This pilot testing resulted in clarification of the wording of existing items and inclusion of additional items. A final survey was developed and reviewed by the project team based on reviewers' comments.

General Information. The survey gathered demographic and background information about the principal and the school in which he or she worked. Items in the school proficiency category consisted of both factual self-report information (such as proficiency on the state assessment) and the principal's perceptions regarding the change in student achievement over time.

School Characteristics. Survey items addressing school characteristics were included to determine if the sample of principals was representative of each state and to

---

[4] In this sense, a measure is a group of items that represent the characteristics in the survey. These items were averaged to create one score on each characteristic.

provide categories by which to disaggregate the data in order to examine how accountability policies might differ by type of district. These items included the number of schools and the number of students in the district, school location (rural, urban, or suburban), per-pupil expenditure, percentage of minority students, and percentage of students receiving free or reduced-price lunch.

Background of Principals. Principals were asked to specify their highest academic degree earned, years of experience in K–12 education in general and as a principal, and length of time in their current principalship. These items were used to determine if there were differences in practices among principals with different backgrounds.

School Proficiency. Each principal was asked to rate the performance level of his or her school with regard to overall proficiency on the state assessment, overall proficiency considering additional sources of data (i.e., dropout rates, expulsion rates, and other assessment results), and consistency in performance across demographic groups (i.e., race, socio-economic status, special education, and English Language Learners). In order to reduce response error, the response sets for these items were very specific with regard to the level of student performance. For example, for the item that asked principals how their school was performing on the state assessment, the response set was 1=most students are below proficient, 3=most students are at proficient and 5=most students are above proficient). Item responses were averaged to create a proficiency measure.

Accountability-specific Measures. For each of the characteristics of effective accountability systems described earlier, measures, or groups of survey items, were developed to ensure the validity and reliability of the results: (1) high expectations for all students; (2) high-quality assessments aligned with standards; (3) alignment of resources, support, and assistance for improvement; (4) applying sanctions and rewards (which also incorporated the multiple measures characteristic[5], (5) district and school personnel data usage; and (6) informative to parents and the community. These measures consisted of items that asked principals about their perceptions related to each of the characteristics of effective accountability systems. The response sets for these items ranged from a low of one to a high of five. Responses for the items were then averaged to create one score on each of the characteristics. The open-ended responses were systematically analyzed for additional detail and contextual information to facilitate interpretations about the measures.

High Expectations for All Students. The items included in this measure were used to examine beliefs about student achievement and overall school performance. The measure assessed the extent of the principal's agreement with five statements regarding staff attitudes and values.

High-quality Assessments Aligned with Standards. Because the state assessment system has become the focal point of accountability under NCLB, survey items were included that measured principals' perceptions regarding the quality of their state

---

[5] The multiple measures characteristic was not reported as an averaged measure because the items were not scaled items but rather asked principals to generate lists of measures that they used and the frequency with which they used them.

assessment. It could be reasoned that if principals considered their state assessment system to be of high quality in terms of alignment to standards or providing usable and useful results, they would be more inclined to incorporate the results in their decision making. Thus, items in this category included how well the state assessment aligned to standards, was informative to parents, provided teachers with access to results, was comprehensive, and provided diagnostic data.

Alignment of Resources, Support, and Assistance for Improvement. This measure was developed to assess whether principals had adequate resources to support changes required by NCLB at the district and classroom levels. This measure included items to identify the adequacy of the resources available to schools. Specifically, principals were asked about the adequacy of available resources to support teachers in using data and to support principals in improving their schools; principals also were queried about the extent to which they had the flexibility to align resources given the school's needs.

Applying Sanctions and Rewards. This measure examined an important aspect of how principals used data in their practices. The five items that made up this measure examined how and to what extent principals incorporated data-driven methods into their practice of monitoring school progress and evaluating staff. This set of items also asked about the degree to which principals thought that the use of rewards and/or sanctions influenced school practices.

District and School Personnel Data Usage. Changes can be made more effectively at the school and classroom levels if school personnel are using data to make informed decisions. Thus, an important consideration for effective accountability systems is not only the ways in which principals are using data to inform decision making but also the ways in which they are supporting their staff in using data. Items in this measure included the extent to which principals encouraged staff to incorporate data into their practices through professional development, established teacher policies and expectations for data usage, and monitored school progress toward data-driven goals.

Informative to Parents and the Community. This measure was used to determine the degree to which principals used data to inform community members, in particular parents, about the progress of schools in the district. In particular, principals were asked about specific formal policies and practices that were in place to ensure parents understood the complexities associated with assessment information.

Reliability analyses were conducted on the measures of the closed-ended items to determine how well individual items within a measure correlated with each other, resulting in the reliability coefficients shown in Table 1. The only measure that showed a low reliability was "Applying Sanctions and Rewards." This low reliability was due to the fact that principals rated one item, "rewards and sanctions influenced their practices," much lower than the others. However, it was determined that this item was an important part of the construct and should not be deleted. Otherwise, the reliability of each of these measures was adequate to warrant using the measures in further analyses and research reporting.

Table 1. Reliability of Survey Measures

Measure	Number of Items in Measure	Sample Size	Reliability
School proficiency	5	116	.83
High expectations for all students	5	118	.88
High-quality assessments aligned with standards	6	118	.91
Alignment of resources, support, and assistance for improvement	6	119	.81
Applying sanctions and rewards	5	97	.41
District and school personnel data usage	13	89	.88
Informative to parents and the community	7	102	.75

SAMPLING

As described in an earlier report (Englert et al., 2003), representatives from the seven state departments of education in the Central Region (Colorado, Kansas, Missouri, Nebraska, North Dakota, South Dakota, and Wyoming) were contacted and asked to participate in a series of three studies. Representatives from four of these states (Colorado, Kansas, Missouri, and South Dakota) agreed to participate. Each state representative was asked to identify a sample of 20 districts that would represent the state on several contextual factors. Specifically, state representatives were asked to identify districts that would represent the state in terms of the percentage of low-, middle-, and high-performing districts; percentage of rural, suburban, and urban districts; varying amounts of per-pupil expenditures; and varying percentages of minority/ELL students[6].Of the 80 districts identified and contacted, 48 superintendents completed surveys regarding their assessment and accountability policies and practices.

The current study is a follow-up of principals in the 48 districts whose superintendents responded to the initial survey. Surveys were sent to a sample of 308 principals in these districts. Emails were sent to 286 of the 308 principals with a link to a website that would allow them to complete the survey electronically. Paper surveys were mailed to 22 principals for whom email addresses were not available. Both electronic and paper surveys included identical questions in order to ensure that the responses would be comparable. Three reminders were sent to the principals. The final number of responses received was 121 (102 responded electronically and 22 responded via mail), resulting in a response rate of 39 percent.

SCHOOL CHARACTERISTICS & PRINCIPAL BACKGROUND

Because the response rate was somewhat low, the characteristics of the responding schools were analyzed in order to determine if they appropriately represented the schools in the states from which they were selected or if sampling bias was a concern. In terms of type of location, approximately 43 percent of principals identified their school as rural (see Table 2). This figure is in the range of statewide averages

---

[6] Having the chief state school officers select the sample may have resulted in sampling bias; however, because the they were most familiar with their states and districts, it was determined that they were best qualified to purposively select appropriate districts.

for the four participating states; according to data available from the Common Core of Data (CCD) from the National Center for Education Statistics (NCES), 31 to 77 percent of schools in the four states are located in rural areas. With regard to socio-economic status, participating principals reported, on average, that 44 percent of students were receiving free or reduced-price lunch. This figure is higher than the CCD reported statewide averages for the four participating states of 28 to 35 percent of students receiving free or reduced-price lunch (see Table 3). In terms of percentage of minority students, principals reported, on average, that 30 percent of students were members of minority groups; this figure is similar to the range of state averages reported by CCD of 14 to 33 percent (see Table 3). In terms of grade level, principals were fairly well distributed across the grades that they served, although higher percentages of principals served elementary schools (58%) than middle and high schools (40%).[7]

These data indicate that the sample is fairly representative of the population of schools in each of the four states. However, higher levels of free or reduced-price lunch participation might skew the results on some of the measures. For example, it is well documented in education research that student proficiency is correlated with student achievement. Therefore, the school proficiency measure might have been slightly higher if the sample was more representative of the state.

Table 2. Participating Schools: Type of Location

School Location	Percentage N=121
Rural	42.6%
Suburban	19.7%
Urban	37.7%

Table 3. School Characteristics & Principal Background

Variable	Mean (N=121)	SD
Number of students in school	540.6	493.2
Percentage of minority students	30.4	26.6
Percentage of students receiving free or reduced-price lunch	43.9	25.4
Years in K–12 education	23.1	8.1
Years as a principal	9.9	7.2
Years as a principal in current district	5.1	4.7

ANALYSES

Several analyses were conducted on the survey data. First, because the primary goal of this study was to gather descriptive data about principals' perceptions of their use of data, means, frequencies, and standard deviations were computed on individual survey items and overall measures. Second, analyses

---

[7] These figures represent principals who selected grade ranges that fell within the defined categories; an additional 2 percent of principals selected multiple categories (such as K–12).

were completed on the open-ended response items associated with each of the measures. Open-ended questions were used throughout the survey to obtain a more comprehensive picture of how principals are using data for accountability purposes and in what contexts. The comments for each item were analyzed and common themes were identified. A theme was considered to be any specific problem, concept, or opinion that two or more respondents included in their written answers to a particular question. The comments were then coded by theme (a single comment could be coded under multiple themes). The open-ended results are discussed in conjunction with the results of the corresponding measures. Schools were then divided into groups based on the information that principals provided about school characteristics, such as rural, urban, and suburban, as well as high and low-SES[8]. Finally, an analysis was conducted on principals who rated themselves higher and lower on the proficiency measure. These analyses revealed the degree to which higher performing schools and those that were improving in terms of achievement were implementing the characteristics of effective accountability systems; these analyses were conducted in order to understand if employing these elements was in fact related to achievement.

FINDINGS

The findings from the analysis of survey data are organized into two sections: accountability-specific measures, and comparisons of high- and low-proficient schools. The findings presented in this section represent the perceptions of surveyed principals in the four states on each of the critical elements described in the literature review.

ACCOUNTABILITY-SPECIFIC MEASURES

The next step in the analysis was to examine the results of the closed- and open-ended items that corresponded to each measure. Table 4 includes the mean and standard deviation for the combined closed-ended items that made up each measure. A 5-point measure was used, ranging from 1=low to 5=high. The following sections describe the results.

Table 4. Descriptive Statistics for Measures

Measure	Mean	SD	Range	Sample Size
School proficiency	3.40	0.59	2.20–5.00	124
High expectations for all students	4.27	0.67	1.80–5.00	119
High-quality assessments aligned with standards	3.59	0.81	1.50–5.00	123
Alignment of resources, support, and assistance for improvement	3.51	0.62	1.83–5.00	124
Applying sanctions and rewards	3.50	0.60	2.00–4.80	124
District and school personnel data usage	3.88	0.51	2.38–4.94	124
Informative to parents and the community	3.44	0.66	1.83–5.00	123

---

[8] High-SES schools are those schools for which the principal indicated that less than 40 percent of students receive free or reduced-price lunch. Low-SES schools are those schools for which the principal indicated that 40 percent or more of students receive free or reduced-price lunch.

School Proficiency

With regard to the overall level of school proficiency, most principals thought that their schools were fairly proficient, rating them 3.40 on average (see Table 4). Comparisons of principal ratings of school proficiency by school locale indicated that, on average, suburban principals ranked their schools' proficiency levels higher (3.58) than urban principals (3.33), yielding a small but meaningful effect size of 0.28.[9] [10]

Principals also were asked to rate the performance of their schools, first when considering multiple sources of data, and second, when considering only performance on the state assessment. Principals rated their schools' proficiency higher in the first case relative to the second (3.15 versus 3.41). This difference may be due to the fact that principals thought that the additional data sources provided a greater scope of information with regard to how students were performing relative to specific goals that were important to the school (attendance, school safety, parental feedback, and district assessments). However, principals generally thought that their schools were improving in their ability to demonstrate proficiency on the state assessment, rating this item a relatively high 3.76.

High Expectations for All Students

This aspect of assessment and accountability systems is intended to contribute to an increase in school effectiveness by creating high expectations for students and ensuring that all members of the school staff value and enforce these expectations. The mean for this measure was 4.27 (see Table 4), and all five of the items in this measure had a similar mean score (4.09–4.38).

High-quality Assessments Aligned with Standards

This measure included items that focused on principals' perceptions concerning the quality and utility of their state assessments, which is critical given the increased focus on state assessments as a primary accountability measure. Overall, the average response to the measure was a fairly high 3.59 (see Table 4). Although principals considered the state assessment to be well aligned to state standards (4.04), their rating of the information it provided to parents (3.45) and the extent to which the assessment provided diagnostic data to inform instructional practices (3.25) was much lower.

Common themes in the open-ended questions were the need for states and districts to provide more diagnostic data at the student level, more timely feedback to help teachers more effectively use data to inform instructional practices, ongoing professional development for school personnel in data analysis and ways in which data have been used to effectively improve instruction, and better technological systems for monitoring data. For example, one principal expressed the need for "a statewide data system that is consistent from district to district." Another principal requested "programming that allows teachers access to all data from their work stations." Allowing teachers "time to discuss [data reports] and reflect

---

[9] The effect size is the difference between the two groups expressed in standardized units. An effect size is the mean of one group minus the mean of a second group, all divided by the average standard deviation. This statistic provides additional evidence when examining group differences. Effect sizes are particularly useful for comparing group differences when a very small or very large sample size is used because the results of significance testing can be unduly influenced by extreme sample sizes.

[10] Effect sizes are not considered to be meaningful if less than .20; those between .20 and .50 are considered small; between .50 and .80, medium; and greater than .80, large (Cohen, 1988).

on best practice" either through early release days or built-in collaboration time was also a district need captured by one principal but expressed by many.

Recognizing the importance of state assessments in school proficiency ratings, improvements in the ways in which data are reported and monitored, and providing ongoing training in proven data-based instructional decision-making strategies may help schools use data more effectively.

Alignment of Resources, Support, and Assistance for Improvement

This measure focused on principals' perceptions of the sufficiency of available resources, their flexibility in determining how to best align those resources, and the level of support they currently have from the district for improving their schools. The mean rating of this measure was 3.51 (see Table 4). Among the individual items included in this measure, two directly related to sufficiency of resources and flexibility to align resources; these received the lowest ratings (2.97 and 3.32, respectively). However, respondents' mean rating of their *ability* to support teachers in using data was 3.39, while their rating of their possession of the *knowledge* and *skills* needed to use data was somewhat higher (3.75). These findings indicate that respondents may have expertise in using data that they are unable to act on due to lack of resources.

Two other items included in this measure inquired about the level of district support. Principals reported receiving a relatively high level of support from their districts (3.84). However, they indicated that the level of support districts provide to teachers in using data was somewhat lower (3.69).

Principals' written comments indicate that schools are lacking the support and resources necessary to fully implement data-driven school improvement. Several issues prevent school staff from achieving a full understanding of the data and then using this feedback to identify appropriate instructional improvements. Lack of time to work with data was mentioned repeatedly as a limitation. One principal listed his or her school's top three issues around data usage as "Time to crunch data. Time to study data. Time to plan curriculum adaptations." Many principals said they would like their districts to provide them with this needed time, a sentiment reflected in the following comment: "[I would like from my district] the time and resources to analyze data. Currently, I have limited time to devote to data analysis and feel that I know how to analyze data but need allowed time…"

Responding principals also reported the desire to receive more training and assistance from their district in analyzing data and identifying instructional strategies. One principal wrote that "[the] staff still does not have a clear understanding of how to read [the data] and what to do with it [sic]." Another would like "a SWAT team for staff development to do staff training of how to translate the data into instructional practices in the classroom."

Respondents wish to see improvements in the means by which their states provide them with student achievement data. A large number of respondents stated that they would like to receive test results from the state in a more timely manner. One explained that "we cannot utilize the information that the tests provide if we do not obtain the results before the last week of school." In addition, principals commented that they need data that are more diagnostic in nature and presented in an easy-to-understand manner. They would like data presented in "multiple formats . . . including by student, teacher, and building" as well as by skill and subject. The "data need to be detailed so teachers can know what specific indicators are strengths and weaknesses."

Many principals pointed out that all of these improvements cost money, and that identifying areas for improvement is for naught if schools do not have the resources needed to implement change. Respondent comments on this issue included the following:

"No funding to help implement the level of expectation for all students."
"[Need] money for programs, remediation, tutoring, collaboration, training, support."
"[Need] the resources to make a difference. Specifically, more staff."
"Continuing to meet goals of NCLB with declining budgets..." *(Named as a top issue in principal's school)*

These findings are important to policymakers who are hoping to increase student proficiency. If adequate funds and support are not available to help schools meet achievement goals, these goals become unattainable.

Applying Sanctions and Rewards

Overall, the average response to the items on this measure was 3.50 (see Table 4). Most principals reported that they specifically used state assessment data to monitor the progress of their school (4.38). They also perceived that monitoring student achievement contributed to the overall success of their school (4.32).

Interestingly, when asked if sanctions and/or rewards were linked to their school's assessment results, half of responding principals (n = 61) said "yes"; the other half (n = 62) said "no." One might expect the majority of respondents to answer "yes" because of NCLB requirements linking student achievement to adequate yearly progress (AYP) and spelling out sanctions for schools for not meeting performance goals. However, this discrepancy might be due to the fact that only a fraction of very high- or very low-performing schools are affected by national- and state-level accountability requirements. This may change in coming years as more and more attention is placed on AYP, which may be more difficult for more schools to meet.

A critical aspect of principals' ability to monitor and make appropriate decisions is the use of multiple data sources. Principals were presented with a list of 18 types of supplemental data and asked to indicate which of these they use to form a more complete "picture" of how their school is doing. Over 80 percent of respondents indicated that they use 5 to 14 of these types of information. The five used by the highest *percentage* of principals are attendance rates (84% of respondents), parent/community feedback (75%), district-created assessments (70%), teacher observations (69%), and scores on standardized tests other than the state assessment (65%). Among high school principals only, 77 percent consider drop-out rates. Principals were then asked to name, of the data sources they use, which three they use most *often*. Respondents indicated that they use district-created assessments most frequently, followed by scores on other standardized tests, miscellaneous assessments (e.g., reading assessments), teacher observations, course grades, and school-created assessments.

Principals of schools in rural, urban, and suburban locales appear to use different numbers and types of supplemental data sources. The average number of data sources used was 9.48 for principals in rural schools, 9.20 for urban, and 8.42 for suburban. This suggests that respondents in rural schools use a wider variety of data sources to assess their schools, while suburban principals reported using a smaller number of sources. In addition, a significantly higher percentage of urban (87%, n=40) and suburban (79%, n=27) principals reported using district-created assessments relative to rural principals (52%, n=19). Of those urban and suburban principals who use district-created assessments, approximately half use them more often than all other data sources. Rural principals do not favor any one source as strongly; their top choice, scores on other standardized tests, is used most often by 34 percent of respondents. These results may be due to the fact that larger school districts have the resources to develop their own assessments and tailor them to their specific needs. Smaller, rural districts may be less likely to develop their own assessments, causing schools in those districts to rely on other data sources. In addition, it is

somewhat easier to use course grades and teacher observations (rural principals' other top choices) with small numbers of teachers and students.

Further comparisons of this construct by school locale revealed that suburban schools (3.30) are less affected by sanctions and rewards than urban schools (3.60) are. These findings resulted in a small but meaningful effect size difference of 0.28, indicating that urban schools were more likely to engage in or be affected by rewards and sanctions.

Principals also were likely to monitor changes and growth in student achievement for their schools using more than one measure. Principals were asked to indicate which of five types of information they use to monitor their schools. Over two-thirds used two or three of these types. The two types used most by respondents were "percentage of students proficient by grade level" (90%) and "AYP" (85%).

District and School Personnel Data Usage

The relatively high average score on this measure (3.88) indicated that most principals are encouraging the use of data as an informative diagnostic tool (see Table 4). Close to 80 percent of the principals surveyed indicated that there was a school policy in place that encouraged teachers to use data to inform instructional decision making, and that focused staff development (4.50) seemed to be an important way in which principals use data to inform their work. Yet, despite a high mean rating (4.34) among principals reporting that data are being used to identify school instructional strengths and weaknesses, many principals expressed a desire for expert guidance or more training on effective identification strategies. For example, one principal stated, "I need district support in summarizing the data in a meaningful way and identifying appropriate teaching strategies which address our weaknesses." Another expressed, "It would be helpful to have data shared and explained that would enable us to change (revise) our instruction." This finding is aligned with relatively low scores on items that asked principals about their perceptions of the availability of school-wide professional development opportunities that helped teachers use data in their classrooms (3.68) and the prevalence of teachers developing a culture of using data to inform classroom practices (3.61).

Principals believed that using data to help improve instructional practices is an important outcome of accountability systems. However, the ways in which data is being used diagnostically varied across schools; but most schools indicated a desire to learn new and more effective ways of translating assessment scores into pedagogical strategies.

Informative to Parents and the Community

Nearly 90 percent (86.3%) of the principals surveyed reported having policies for communicating accountability results to the community; a high mean rating (4.32) indicated that they strongly believed that teachers were implementing these policies (see Table 4). Principals also reported disseminating up-to-date accountability information in multiple ways (3.54) but rated translation of these materials for non-English-speaking parents lower (2.96).

When asked what kinds of assessment data are reported to the public and how, many principals indicated in open-ended items that state standardized tests and district assessments are communicated through parent-teacher conferences, progress reports, mailings, newsletters, and parent nights. A mean rating of 3.46 further indicated that principals perceived they are reporting adequate amounts of information to the community about what they are doing to meet accountability demands.

COMPARISONS OF HIGH- AND LOW-PROFICIENT SCHOOLS

As mentioned earlier, principals were asked to rate the degree of proficiency of their schools considering several variables including overall proficiency on the state assessment, improvement on the state assessment, degree of differences in achievement among subgroups, and performance on other measures of achievement. Schools were divided into two groups, one scoring below the mean on the measure and the other scoring at or above the mean.[11] These contrasting groups of schools were then compared using effect sizes and t-tests based on the degree to which they differed on the characteristics of effective assessment and accountability systems.

These analyses revealed substantial differences between schools whose principals rated their school proficiency higher and those whose principals reported lower school proficiency ratings (see Table 5). Most notably, principals who rated their schools highly proficient had higher expectations for their students, sufficient resources and support for improvement, used data diagnostically, and thought their state assessment was of high quality. All of these measures were statistically significant and had small to medium effect sizes. The only two measures that were not significantly different across groups were the degree to which sanctions and rewards were being used (although a small effect size difference existed) and communicating results to parents and the community. These results show that principals of those higher performing schools rate their employment of most (4 out of 7) of the elements of effective assessment and accountability systems more highly than principals in those schools rated as lower performing.

Table 5. Comparisons of Measure Scores for Low- and High-Proficient Schools

	Low Proficient N=57		High Proficient N=61		Effect Size
	Mean	**SD**	**Mean**	**SD**	
High expectations for all students	4.06	0.76	4.49	0.48	0.69**
High-quality assessments aligned with standards	3.41	0.89	3.80	0.70	0.49*
Alignment of resources, support, and assistance for improvement	3.33	0.64	3.71	0.54	0.64**
Applying sanctions and rewards	3.42	0.63	3.58	0.56	0.27
District and school personnel data usage	3.80	0.55	3.97	0.45	0.35*
Informative to parents and the community	3.39	0.66	3.45	0.64	0.09

* Significant to p<0.05. **Significant to p<0.01.

One logical question that arises from these analyses is the difference in achievement related to the schools' population (i.e., higher percentages of low-SES students or higher percentages of minorities). Not surprisingly, slight differences were found between the two groups. On average, 22 percent of students in schools in the high-performing category were members of minority groups compared to 40 percent in the low-performing category. In addition, 38 percent of students in high-performing schools were receiving free or reduced-price lunch compared to 50 percent in the low-performing category. These

---

[11] Although it may be preferable to split the groups into the top and bottom third of the distribution to produce groups that are more distinct, the sample was too small for this type of analysis.

findings make it difficult to discern if the differences on the measures were related to the higher levels of performance or if these elements are confounded by the differences in background variables such as SES.

Another area of interest was the degree to which schools are employing elements of effective assessment and accountability systems and the degree to which principals reported improvements in their school-wide achievement. One item on the survey asked principals to report the degree to which their school was improving on the state assessment. For the purpose of comparison, schools were divided into two groups: one group consisted of principals who rated their school achievement as declining or static; the other group consisted of principals who indicated that their school was improving or significantly improving.

The analyses showed that schools whose principals rated them as improving were much more likely to use most of the elements of effective accountability systems (see Table 6). In particular, principals of improving schools were much more likely to think that they had adequate resources to support the changes required by NCLB and to improve practices at the school and classroom level. The differences between schools that were improving on the state assessment and those that were either stagnant or declining was statistically significant (p<0.01), with a large effect size. It seems reasonable to assume that this type of support would be critical to schools because changing education practices, such as implementing new curricula and delivering quality professional development, requires a monetary and time investment.

Table 6. Comparisons of Measure Scores — Schools Declining or Static on State Assessment & Schools Improving on State Assessment

Measure	Declining or Static Achievement Levels* N=34		Improving Achievement Levels* N=75		Effect Size
	Mean	SD	Mean	SD	
High expectations for all students	4.09	0.81	4.36	0.57	0.39*
High-quality assessments aligned with standards	3.49	0.97	3.66	0.74	0.20
Alignment of resources, support, and assistance for improvement	3.19	0.57	3.68	0.59	0.84**
Applying sanctions and rewards	3.44	0.55	3.53	0.62	0.15
District and school personnel data usage	3.71	0.50	3.96	0.50	0.50**
Informative to parents and the community	3.27	0.59	3.49	0.67	0.35*

* Significant to p<0.05. **Significant to p<0.01.
Note: Based on principals' reports

It seems reasonable that if teachers and principals are using data in the ways described earlier, they might be better positioned to make informed decisions about allocating and aligning resources to reach those goals. Thus, it is not surprising that principals of improving schools are using data to inform their decisions and to support teachers in using data in the classroom. The difference between the two groups of schools on this measure (school personnel data usage) was statistically significant and resulted in a medium effect size.

Three other measures — high expectations, informing parents, and high-quality assessments — showed small effect size differences; two were statistically significant. These measures indicate that schools that are improving their achievement are holding all students to high expectations and inviting parents and community members to be more involved in the educational process by providing accountability information. It is also logical that schools that were improving on the state assessment would value the test and work towards improvement.

Interestingly, the "applying sanctions and rewards" measure was neither statistically significant nor was the effect size meaningful. One explanation for this finding may be that schools that are working to increase achievement may not be focused on the punitive aspects of accountability (i.e., they are doing well so there is no need for sanctions). However, this difference might also be due to the low reliability of the measure discussed earlier which may impact the sensitivity of this measure in detecting differences between the two groups.

In order to further understand these findings, the demographics of the two categories of schools were examined. On many school background questions, such as principals' years of experience and student enrollment, no differences were found between the two groups. However, differences were found in the student composition of the two groups of schools. In schools whose principals reported that student achievement on the state assessment was increasing, 31 percent of the students were minority and 46 percent were receiving free or reduced price-lunch compared to 28 percent and 38 percent, respectively, of the students in schools where achievement was declining or static.

One might think that these demographic differences may have some impact on the interpretation of the measure differences mentioned earlier. It may be that schools with higher levels of minority students and students receiving free or reduced-price lunch had more room to improve on state assessments when compared to more affluent districts that might already be high achieving. However, after examining frequencies on these categories, this does not appear to be an issue. The majority of principals who rated their school-level achievement as proficient also rated their school's achievement as improving (42.7 %). This indicates that principals of these schools do not think they have reached a limit on their capacity to improve on the state assessment. This result is quite high when compared to 6.5 percent of the principals in proficient schools who believed they were not improving. However the differences in the results between Tables 5 and 6 are due to a group of principals who said their schools were in the "not proficient" category in Table 5 but in the "improving" category in Table 6. Specifically, this group, which accounts for almost one quarter (23.4%) of the respondents, is in the group who were in a lower category in Table 5 and then in a higher category in Table 6. It is important to note that, relative to the overall sample, these schools are more likely to be urban (50%) and have higher percentages of minority students (48%) and students receiving free or reduced-price lunch (58%). This finding indicates that these practices may be particularly relevant and important to schools with high percentages of at-risk populations.

CONCLUSIONS AND NEXT STEPS

This study helps to illuminate how a group of principals use data in light of new state and national requirements for accountability. On a very positive note, the responding principals reported using data to a great extent in their own practices, and they are encouraging their staff to use data in ways that align with the notion of effective accountability practices in accordance with many of the policies identified in the research literature (i.e., the seven characteristics of effective assessment and accountability practices). Ideally, following these practices will result in improved pedagogy and help inform curricular decisions. For example, principals overwhelmingly reported holding all of their students to high standards and ensuring that staff understand and support this philosophy. Many principals also expressed an interest in

learning the most effective strategies for using accountability data in ways that will improve student learning and achievement. In addition, principals cited a lack of time as a major limitation in their ability to effectively use data and to implement change. A majority of principals reported that a lack of time and limited access to other resources (e.g., collaborative planning time, time to learn effective assessment and diagnostic practices, instructional specialists well versed in using data informatively, money) inhibited them from maximizing their effectiveness in supporting teachers to use data in diagnostic ways.

It is difficult to ascertain the effectiveness of sanctions in improving student learning. Differences in the responses between urban and suburban principals may indicate that suburban schools have a stronger capacity to meet NCLB student proficiency demands. More research is needed in this area, however, before conclusive arguments may be made.

One of the important findings of this study was the survey score differences that indicated that responding principals of high-performing or improving schools were more likely to use the elements of effective accountability systems identified in the literature review. Effect size differences for both groups of schools showed that they reported using these accountability practices much more often on all but one of the six measures. Particularly noteworthy is the fact that these two groups of schools (especially the improving schools) reported using these elements; this was true even for schools with relatively large proportions of minority and low-SES students.[12]

Though the results from this study provide useful information about self-reported principal practices, additional data are needed to obtain a more complete picture of the impact of accountability at the school and classroom levels. In subsequent months, the teachers of principals who participated in this study will be surveyed to examine the interrelationships between self-reported district, school, and classroom practices and policies. This nested data structure will allow for a direct analysis among these three levels.

---

[12] Although this study provides descriptive findings of what is currently happening in schools regarding critical assessment and accountability issues, it is important to remember that the findings are in no way causal.

REFERENCES

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

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

Buckendahl, C. W., Impara, J. C., & Plake, B. S. (2002). District accountability without a state assessment: A proposed model. *Educational Measurement: Issues and Practices, 21*(4), 6–16.

Cohen, J. (1988). *Statistical power analysis in the social sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.

Creemers, B. P. M. (1994). *The effective classroom.* New York: Cassell.

Education Week. (2002, January 10). Quality Counts 2002: Building blocks for success [special report]. *Education Week, 21*.

Englert, K., Fries, D., Goodwin, B., & Martin-Glenn, M. (2003). *Understanding how superintendents use data in a new environment of accountability.* Aurora, CO: Mid-continent Research for Education and Learning.

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

Goodwin, B., Englert, K., & Cicchinelli, L. F. (2003). *Comprehensive accountability systems: A framework for evaluation* (REL Deliverable #2002-07). Aurora, CO: Mid-continent Research for Education and Learning.

Grissmer, D., & Flanagan, A. (1998). *Exploring rapid achievement gains in North Carolina and Texas.* Washington, DC: National Education Goals Panel.

Grissmer, D., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What state NAEP test scores tell us.* Santa Monica, CA: RAND.

Guth, G. J. A., Holtzman, D. J., Schneider, S. A., Carlos, L., Smith, J. R., Hayward, G. C., et al. (1999). *Evaluation of California's standards-based accountability system.* Menlo Park, CA: WestEd, Management Analysis and Planning Inc.

Marzano, R. J. (2000). *A quantitative synthesis of research on school-level, teacher-level, and student-level variables related to academic achievement* [REL Deliverable 2000-05]. Aurora, CO: Mid-continent Research for Education and Learning.

The Princeton Review. (2002). *Testing the testers 2002: An annual ranking of state accountability systems.* New York: Author.

Reeves, D. B. (2002, March/April). *Accountability-based reforms should lead to better teaching and learning - period.* Retrieved October 27, 2003, from www.edletter.org/current/reeves.shtml

Siegel, D., & Fruchter, N. (2002). *Final report: Evaluation of the performance-driven budgeting initiative of the New York City Board of Education*. New York: New York University, Institute for Education and Social Policy.

Sirotnik, K. A., & Kimball, K. (1999). Standards for standards-based accountability systems. *Phi Delta Kappan, 81*(3), 209–214.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research.* New York: Falmer Press.

Torrence, V. D. (2002). *Principals use of data: A national perspective.* Doctoral dissertation, Virginia Polytechnic Institute. Retrieved October 7, 2004, from http://scholar.lib.vt.edu/theses/available/etd-04202002-124945/unrestricted/DissertationVer10.pdf

Trevisan, M. (1999). Administrator certification requirements for student assessment competence. *Applied Measurement in Education, 12*(1), 1–11.

Walberg, H. J. (2002). Principles for accountability designs. In H. J. Walberg (Ed.), *School accountability* (pp. 155–183). Stanford, CA: Hoover Institution Press, Stanford University Press.