HOW ARE EDUCATORS USING DATA?

A COMPARATIVE ANALYSIS OF SUPERINTENDENT, PRINCIPAL, AND TEACHERS' PERCEPTIONS OF ACCOUNTABILITY SYSTEMS

Regional Educational Laboratory
Contract #ED-01-CO-0006
Deliverable #2005-07

Prepared by
Kerry Englert, Ph.D.
Dawn Fries
Mya Martin-Glenn, MA
Sheri Michael, MA

November 17, 2005

This document has been funded at least in part with federal funds from the U.S. Department of Education under contract number ED-01-CO-0006. The content of this publication does not necessarily reflect the views or policies of the Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

TABLE OF CONTENTS

INTRODUCTION

This report evaluates the findings from three previous research studies that examined how different groups of educators perceive the effectiveness and usefulness of their accountability systems. The information presented in these studies focuses on understanding how educators incorporate effective assessment and accountability practices into their work, as well as examining how they view the quality of the associated assessments and policies. 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 second study expanded that work by surveying principals in the same districts in order to understand how assessment and accountability practices are actualized in schools (see Englert, Fries, Goodwin, Martin-Glenn & Michael 2004). The final study in this series examined teacher responses to these policies and practices (see Englert, Fries, Martin-Glenn & Michael 2005). This report examines these findings in total by considering the consistencies and cohesiveness across superintendent, principal and teacher perceptions of their accountability systems, and examines how these three groups utilize information differently in view of the levels of student achievement in their school or district.

The purpose of this study is to develop a better understanding of the assessment and accountability practices and policies that educators are implementing in the classrooms, schools and districts and to examine whether those policies are associated with perceived school and student improvements in achievement. This study generated descriptive information about the need for schools and districts to effectively use data, how schools and districts use data to guide classroom practice, and the differences in data use based on the level of student proficiency in individual schools.

This report serves as the technical report on the combined study findings and is intended to further the understanding among educators and policymakers regarding issues of accountability and how policies might be interpreted differently by stakeholder groups. Future reports related to this work will include articles directed to state department of education personnel on data use in districts and schools and practices that tend to be associated with improving high achieving schools. Additionally, another publication will be directed to school and district administrators to help them examine their current practices and how these practices relate to effective assessment and accountability practices.

SEVEN CHARACTERISTICS OF EFFECTIVE ACCOUNTABILITY SYSTEMS

> The logic behind the No Child Left Behind Act is that by testing students frequently and holding teachers and administrators accountable for achievement test results, student learning will improve. The law assumes that, by examining annual achievement data, educators can determine what causes unacceptable outcomes and can correct the unproductive parts of the system. But how can *processes* be improved by inspecting only their *outcomes*? (Heibert, Stigler, Jacobs, Givvin, Garnier, Smith, Hollinsworth, Manaster, Wearne, Gallimore, 2005, p112)

The effectiveness of schools' accountability policies is of paramount importance to the success of the No Child Left Behind Act (NCLB). If schools are not actively engaged in effectively using

accountability data, the student achievement increases required by this legislation will likely be unattainable. Many education researchers and scholars cite both positive and negative impacts of this change in emphasis. 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. The negative impacts of increased accountability cited by researchers include 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]. Many of these components are relevant to the way educators use and perceive their accountability systems, including (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) ue of multiple measures; (6) diagnostic uses for data; and (7) readily understandable to the public. These elements (or a subset thereof) were selected and used as the foundation for constructing three surveys each of which were given to a superintendents, principals, and teachers. Grounding the research study in these characteristics encourages the use of theories discussed in the literature to determine if these components are related to improvements in student achievement. 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 that 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, advantaged or disadvantaged, meet state standards.

One way to ensure that school staff members focus attention on all students is to test and monitor students and subgroups of students. By doing so, 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 suggests that maintaining and communicating a school-wide press for academic achievement is a contributing factor to helping low-achieving students meet higher standards, consequently creating an effective school (Teddlie & Reynolds, 2000; Creemers, 1994; Marzano, 2000). This notion of

[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.

high expectations is a cornerstone of the NCLB legislation, as 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. Assessment data are the primary source of information under NCLB, and 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 clear academic standards are specified by the state and district, student progress on standards needs to be measured using assessments that are aligned to those standards (Guth, Holtzman, Schneider, Carlos, Smith, Hayward, et al., 1999; Baker, Linn, Herman, Koretz, 2002; Buckendahl, Impara, & Plake, 2002). Moreover, most accountability standards and models also 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 relevancy of the 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 provide 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, 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 increases 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 include the criterion of whether the state "holds schools accountable for performance" (p.12) through rewards and sanctions, including closure, reconstitution, transfers, and withholding of funds. And most notably, NCLB holds schools and districts accountable for increases in student achievement. If gains are not made according to a predetermined rate, schools and districts face sanctions.

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 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 (Baker et al., 2002; Sirtonik & Kimball, 1999). 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.

Data Use in Districts and Schools[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 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. A study by Fuhrman (1999) supporting the importance of data usage found that 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.

Informs Parents and the Community[3]

Another commonly identified element of effective accountability systems is the capacity to inform the parents and community 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 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

---

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

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

knowledge. It would be most useful 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 for members of the public to be provided data combined 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 listed above relate specifically to issues that affect administrators and teachers (such as how they perceive the quality of the state assessment) or issues that they have the ability to implement or affect (such as using data to inform their work). NCLB has increasingly focused the education community on assessment and accountability. Past studies have shown that principals, superintendents and teachers are increasingly using data in their work (Englert, Fries, Goodwin, Martin-Glenn, 2003; Englert, Fries, Goodwin, Martin-Glenn, Michael 2004; & Englert, Fries, Martin-Glenn, Michael, 2005). This study expands on the previous studies by examining how perceptions across these groups of educators are consistent or disparate. For example, are superintendents more likely to use data to inform their work than are teachers? Information from this study presents a view of how these groups of educators are functioning in this era of accountability. The background literature presented focuses this research study on the areas that are directly relevant to the *effective* use of accountability data. The results are intended to provide information that could be used to facilitate a conversation around meaningful issues that can directly impact and improve substantive areas to influence constructive uses of data.

Research Questions

In order to understand the consistencies in the use and perception of assessment and accountability systems, the following research questions guided this study:

- How similar are superintendents', principals', and teachers' beliefs regarding policies that are being implemented to meet new accountability demands?
- Are administrators and teachers consistent in how they implement policies and practices that research has identified as being effective?
- Do policies and practices of principals and teachers in high-performing schools differ from those in low-performing schools?

Method

In order to provide answers to these three research questions, McREL conducted a series of three surveys. The surveys measured the degree to which superintendents, administrators, and teachers incorporated the effective characteristics described above into their practices, as well as how they perceived school and district policies and other assessment related issues. The section

that follows describes the development of the survey, including the incorporation of the different characteristics of effective accountability systems.

Instrument Development

McREL developed each of the three surveys based on the relevant elements derived from the literature, as well as district, school and educator background information. Closed- and open-ended items were developed that aligned with each of the different characteristics. Each survey measured educator's 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 being employed in their classrooms, schools, and districts. Measures[4] were developed from the closed-ended items to provide more consistent indicators of the constructs.

Prior to distribution, each survey was pilot tested with a small group of educators who were either current or former teachers or former administrators. This pilot testing resulted in clarification of the wording of existing items and inclusion of additional items. Final surveys were developed and reviewed by the project team based on reviewers' comments.

General Information. Each survey included demographic and background questions about the respondents and the school or district in which they worked. Items in the student proficiency category consisted of both factual self-report information (such as proficiency on the state assessment) and perceptions regarding the change in student achievement over time.

School/District Characteristics. Survey items addressing these characteristics were included to determine if the sample of educators was representative of each state and to provide categories by which to disaggregate the data in order to examine how accountability policies might differ by type of school. Superintendents were asked questions about their districts, and principals and teachers were asked questions about their schools. This category included items such as the number of schools in the district, the number of students in the school, school/district location (rural, urban, or suburban), per-pupil expenditure, percentage of minority students, and percentage of students receiving free or reduced-price lunch.

Background of the Educator. Each respondent was asked to specify their highest academic degree earned, years of experience in K–12 education, and length of time in their current position.

Student Proficiency. Each respondent was asked to rate the performance of the students in his or her district, school, or classroom 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

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

were very specific with regard to the level of student performance. For example, for the item querying how students were performing on the state assessment, the response set was 1=most students are below proficient, 3=most students are at proficiency and 5=most students are above proficiency). Item responses were averaged to create a measure of overall proficiency.

Accountability-specific Measures. For each of the characteristics of effective accountability systems, measures, or groups of survey items, were developed to provide more accurate information about the characteristic: (1) holding high expectations for all students; (2) utilizing high-quality assessments aligned with standards; (3) sufficiency of the alignment of resources, support, and assistance for improvement; (4) applying sanctions and rewards based on results (which also incorporated the multiple measures characteristic[5]), (5) using data to inform practice; and (6) informing parents and the community about results. These measures consisted of items that asked the educators about their perceptions related to each of the characteristics of effective accountability systems. The response options 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 teacher or principals'[6] 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 respondents' perceptions about the quality of state assessment. It could be reasoned that if they 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 each respondent felt they had adequate resources to support changes required by NCLB. This measure included items to identify the adequacy of the resources available to them.

---

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

[6] Only principals and teachers were asked questions about high expectations. This was done because it was felt that superintendents might be too far removed from the actual student to make a judgment about this and to keep the survey as short as possible.

Specifically, they were asked about the adequacy of available resources to support data usage and to support school improvement efforts.

Applying Sanctions and Rewards. Because much of the focus on NCLB has been around using sanctions and rewards as a means to increase student achievement, these items asked to what degree each respondent felt this was prevalent in his or her own school/district. For example, this measure examined the degree to which the educator was required to meet data-driven performance goals and to the degree that they were evaluated based on changes in student achievement.

District and School Personnel Data Usage. Changes can be made more effectively at the district, building, or classroom level if educators are utilizing data to make informed decisions. Items in this measure included the extent to which educators incorporate data into their practices such as identifying instructional strengths and weaknesses, aligning curriculum to state standards, identifying instructional strengths and weaknesses, and focusing on professional development.

Informative to Parents and the Community. This measure was used to determine the degree to which teachers used data to inform community members, in particular parents, about the progress of their students. In particular, teachers were asked about specific formal policies and practices that were in place to ensure parents understood the complexities associated with assessment information, the ways schools or teachers communicated results to parents, and the ways parents could support student learning at home.

Reliability analyses were conducted on the measures of the closed-ended items to determine how consistent the respondents' answers were to the set of items resulting in the reliability coefficients shown in Table 1. The coefficients ranged from 0.41 to 0.89. There were several measures that were a concern because they had low reliabilities (0.41 and 0.49). These scales were examined to determine if specific items were contributing to the low reliabilities. One measure that showed a low reliability was "Applying Sanctions and Rewards." This low reliability was partially due to the fact that principals' responses for one item, "rewards and sanctions influenced their practices," were inconsistent with how they rated other items. However, further analysis indicated that this item was an important part of the construct and should not be deleted. With regard to the teacher responses on this measure, the reliability was again low. Contrary to the principal responses, there was not a single item that contributed to the low reliability. The low reliability instead might be due to the fact that teachers are not conceptualizing or focusing on the issues of "rewards and sanctions" in the same way. For instance, the superintendents' responses resulted in an adequate degree of reliability and this may be because they are focused on these issues more frequently and can respond in a much more consistent manner. The other reliability results indicate that these measures had adequate reliability to warrant using them in further analyses and research reporting. However, consideration was given to the measures with lower reliabilities when discussing the findings from the analyses of those measures.

Table 1. Reliability of Survey Measures

Measure		Superintendents		Principals		Teachers	
	Number of Items in Measure	*Sample Size*	*Reliability*	*Sample Size*	*Reliability*	*Sample Size*	*Reliability*
High expectations for all students	5	NA	NA	118	0.88	151	0.84
High-quality assessments aligned with standards	6	47	0.87	118	0.91	148	0.89
Alignment of resources, support, and assistance for improvement	5	49	0.82	119	0.81	151	0.79
Applying sanctions and rewards	5	48	0.76	97	0.41	87	0.49
District and School Personnel Data usage	10	34	0.74	89	0.88	85	0.85
Informative to parents and the community	7	41	0.58	102	0.75	85	0.76

SAMPLING

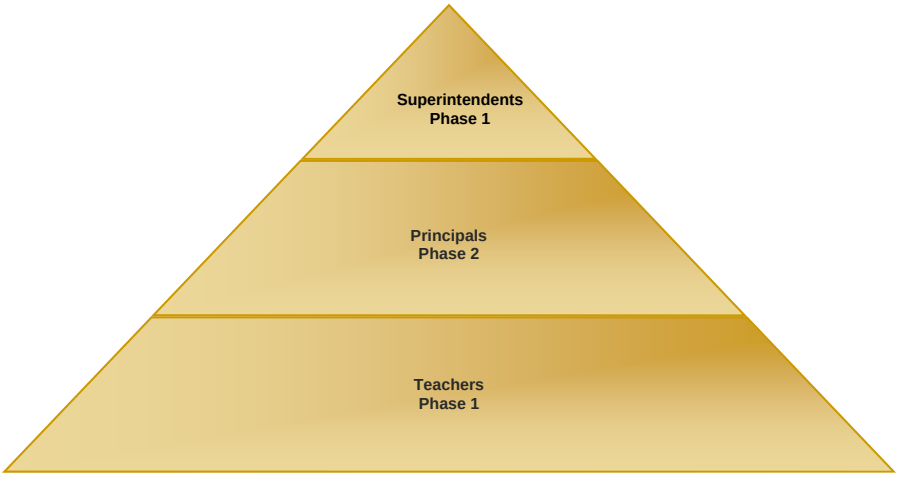
As described in earlier reports (Englert et al., 2003, 2004, 2005), contacts from the seven state departments of education in the Central Region (Colorado, Kansas, Missouri, Nebraska, North Dakota, South Dakota, and Wyoming) were contacted and asked to participate in the series of three studies. State departments from four of these states (Colorado, Kansas, Missouri, and South Dakota) agreed to participate. Each contact was asked to identify a sample of 20 districts that would be representative of the state on several contextual factors. Specifically, they 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[7]. Of the 80 districts identified and contacted, 49 superintendents completed surveys regarding their assessment and accountability policies and practices. For the study on principals, a sample of administrators in the districts whose superintendents participated was selected to participate; this resulted in 121 principals responding. For the final sample of teachers, the responding principals were contacted to see if their teachers could participate. Of those contacted, 25 schools participated, resulting in a sample of 153 teachers completing surveys[8] (see Figure 1). For some of the analyses, the entire sample from each population was used. For other analyses, a sample

[7] 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.

[8] For more detail on sampling issues see individual reports (Englert et. al. 2003, 2004, 2005).

was pulled from the larger sample that represented all of the teachers (N=153), only the principals of those teachers (N=27), and then the superintendents of those principals (N=19). The latter allowed for a comparison of how the different groups of educators responded relative to each other within a system while the analyses from the complete sample allowed for results that were more general across the groups.

Figure 1: Phases of Data Collection



Although the sample size was too small to conduct hierarchical analyses, consideration was given to the fact that teachers were nested in schools and how this might impact the subsequent analyses. For example, it is possible that larger schools could contribute more teachers to the sample, potentially skewing the results to reflect the perceptions of teachers in larger schools. Upon examination of the response rate, McREL concluded that this is unlikely. On average, six teachers responded per school. There were three cases where schools had 10 or more teachers responding. The average enrollment for these three schools was 514. The average enrollment for all of the schools with teachers responding was 475, which was not substantially different.

Surveys were administered to superintendents, principals, and teachers over the course of two school years. Superintendents were surveyed in the fall of the 2003/2004 school year, principals in the spring of the same school year, and teachers in the winter of the 2004/2005 school year. The staggered rate at which they were sampled might lead to differences in their responses and any possible impact will be discussed in the conclusion section.

SCHOOL CHARACTERISTICS & BACKGROUND INFORMATION

Overall, the sample sizes were somewhat small for each of the three groups. Therefore, the characteristics of the respondents were analyzed in order to determine the degree to which they appropriately represented schools and districts teachers in the states from which they were selected or if sampling bias was a concern. In terms of type of location, approximately 63, 42, and 51 percent of superintendents, principals, and teachers, respectively, identified their schools as rural. According to data available from the Common Core of Data (CCD), which show that 31 to 77 percent of schools in the four states are located in rural areas, this figure is within the range of statewide averages for the four participating states. With regard to socio-economic status, respondents reported that approximately 44 percent of students were receiving free or

reduced-price lunch. This figure is higher than the CCD-reported statewide averages of 28 to 35 percent free- and reduced-lunch recipients for the four participating states. In terms of grade level, principals and teachers were fairly well distributed across the grades that they served, but there were slightly higher numbers of respondents from elementary and middle schools (for more detail on the sample characteristics see Englert et. al. 2003, 2004, 2005).

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. Consideration of this difference may impact the conclusions that could be drawn from the results.

ANALYSES

The first set of analyses examined mean differences on the constructs between the groups of respondents (superintendents, principals, and teachers) using the entire sample from each group. These results show interesting differences in the manner and degree to which the different groups of educators perceive and are implementing the different components of accountability systems, such as the quality of their assessment systems and how the data are being incorporated into the decision making process. Perhaps not surprisingly, of the three groups, teachers rated many of the different components of effective accountability systems the lowest. (See Table 2).

Table 2. Overall Means and Effect Sizes for Respondent Groups

Variable	Super-intendent Mean (SD) N=49	Principal Mean (SD) N=124	Effect Size (Super-intendent – Principal)	Teacher Mean (SD) N=153	Effect Size (Principal – Teacher)
High expectations	NA	4.27 (0.67)	NA	4.19 (0.59)	0.13
High-quality assessments	3.46 (0.63)	3.59 (0.81)	-0.18	3.29 (0.73)	*0.39***
Alignment of resources, support, and assistance for improvement	3.12 (0.77)	3.51 (0.62)	**-0.56****	3.48 (0.73)	0.04
Applying sanctions and rewards	3.96 (0.59)	3.50 (0.60)	**0.77****	3.22 (0.79)	*0.40***
District and school personnel data usage	3.66 (0.59)	3.88 (0.51)	*-0.40*	3.44 (0.63)	**0.77****
Informing parents and community	4.07 (0.55)	3.44 (0.66)	***1.04****	3.20 (0.80)	*0.33***

Significant to p<0.01. Effect sizes in *Italics=Small,* **Bold=Medium, ***Large=Bold Italics***[9] [10]

[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 pooled 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 generally considered to be "small" if between .20 and .50, "medium" if between .50 and .80, and "large" if greater than .80 (Cohen, 1988).

The component that is perhaps the most critical in terms of implementing an accountability system is ensuring that the data are used to make better and more systematic decisions. This is one of the intentions of NCLB — that by monitoring student achievement, superintendents, principals, and teachers will be able to make the necessary instructional and organizational improvements to address and correct any weaknesses. This goal might only be met if data are used systematically across different levels in the educational system from administration to the classroom. The data in this study were analyzed to examine those consistencies and the results indicate that overall, teachers rated their data use much lower than principals and superintendents. Mean differences between teachers and principals resulted in statistically significant and medium effect size differences.

Respondent comments suggest several reasons why teachers rated their level of data use lower than school and district administrators. The most pressing reason, noted by both principals and teachers, appears to be a lack of time. While principals and superintendents also objected to not having enough time to critically examine data, comments indicated that this need is more pronounced for teachers. One principal explained that "...there is so much pressure on teachers to cover the curriculum and get kids ready for state tests that it's difficult for teachers to find time to be reflective about their teaching practices...." The following quote from a teacher illustrates their perspective:

> I would like to have the time, during the teaching day (not before school, at lunch, after school, or during summer vacation), to evaluate/make sense of data....Teachers at my school are swamped with many duties, and with internal and external pressures to change or adapt curriculum, strategies, and instructional delivery styles to meet NCLB standards, with little time or training given to us so that we may determine the best way to do this....

Principals also noted that their teachers need more training on data use and data-driven instructional strategies. "We need…staff development to help us see how we can best use the data to improve instruction. We're getting better at gathering data, but often I feel teachers don't know what to do with it once they have it...." wrote a principal. Some principals would like more training in data analysis for themselves as well.

In addition to these needs, when asked what their top issues were around using achievement data, many teachers voiced complaints about the assessment system in general. Several teachers responding to the survey noted that their state's assessments do not accurately reflect students' knowledge and skills. One teacher stated emphatically that the state assessment "does not show what the students can do." Another commented that "achievement data does not give the whole picture — there are many components to students' achievement that cannot/are not addressed (i.e., home life, attitude, intellectual ability)." Others expressed disagreement with the whole system of high-stakes testing in general. One teacher wrote that "I don't want to use it [the state assessment]. I would like to focus on my class in my community without expectations from NCLB." Teachers' lack of time and training may explain why they work with data less than the other respondent groups.

Similar differences between teachers and administrators were also found when examining how these educators view their needs around what resources they had available to help them use data. When asked to list such tools or resources, superintendents and principals frequently listed computer software programs and dedicated personnel. However, when teachers responded to this question, they most frequently cited the test data itself. This is noteworthy because it indicates that teachers may not have access to resources that would allow them to efficiently use data. For example, if they don't have adequate hardware or software, manipulating data to

address questions they might have about student achievement data become cumbersome if not impossible.

In another question, respondents were asked to describe the policy or expectation in their school that teachers use assessment data to inform their classroom practices. Overwhelmingly, the administrators indicated that teachers are expected to plan their instruction and/or curriculum for their students based on the needs reflected in the data. One superintendent stated "we expect that they do this and hold them accountable for it. We have also provided training for them." A principal echoed this statement by saying "our staff spends a great deal of time at inservices going over data, sharing ideas on how to improve instruction, and how to significantly reach every student". However, teachers' responses did not reflect a high degree of dialogue between administrators and teachers or that training was provided to help them address weaknesses in student achievement. One comment that was indicative of teachers' attitudes was "it is just asked of us and we do it."

Educators also differed on the adequacy with which parents and community members are being informed of their student's and/or school's progress. While superintendents rated relatively highly the extent to which parents are informed of assessment results, principals' mean rating was significantly lower, and teachers' was lower still. When asked to comment on policies regarding sharing assessment information with parents, superintendents tended to simply state that a policy was in place, without giving many details as to its implementation. "Informing parents in writing" and "We follow state requirements on this" were typical of many responses. Principals and teachers tended to be more specific as to how student results are shared with parents. The most often cited method was direct teacher-parent communication at conferences and meetings; responses indicate that parents are usually given written information in the form of a letter or report card as well. This trend to share assessment results in person could be seen as an opportunity for school staff to have dialogues with parents about the results and to work with them to ensure their understanding of the data. However, principals' and teachers' descriptions of how data are shared were cursory (e.g., "results are communicated," "scores are given," "results are explained"). Neither group of respondents mentioned any established techniques or guidelines in practice to assist teachers in ensuring that parents actually understand assessment results, or that strategies were provided to parents on how to best support their child's learning at home. Moreover, the above finding that teacher data use is limited due to lack of time, knowledge, and personal investment suggest that teachers themselves may not have a full understanding of assessment data. Schools might not be taking steps to provide information through creative channels to reach parents who may be difficult to reach, or those who have different language or cultural backgrounds. Communication of student and school success becomes a critical issue when examining the role of accountability data, particularly as parents face decisions such as evaluating school choice options or supporting their child's learning at home. Schools and districts need to examine their policies around providing accountability data in order to be responsive and responsible to the needs of the public for information about their schools.

Differences between educator groups were also apparent when it came to the application of sanctions and rewards. However, any examination of the results should consider that this measure had a somewhat low reliability among teacher and principal groups, which influences the precision at which these are measured; thus, the results should be viewed somewhat cautiously. As compared to principals, superintendents were more likely to use data to monitor progress of schools towards data-driven goals and to attach consequences to those results. Interestingly, teachers were even less likely than principals to respond that rewards and sanctions were a factor at their school. These results could mean that superintendents are in a position

where they are more attuned to the issues of sanctions because of NCLB and that information has not filtered down to the building or classroom level.

Differences were also apparent when it came to issues of resources. Superintendents rated this component much lower than either principals or teachers. This result was surprising, given the oft held assumption that teachers working in the trenches might feel the most pressure from not having enough supplies, computers, and curricular materials. These results, however, might be due to the fact that district administrators are more focused on issues of budgets and resources because they are responsible on a day-to-day basis for directing resource allocations making this issue more salient to them. Analysis of Responses from Educators Based on Improvements in Student Achievement

While the findings presented above provide a picture of the overall consistency between groups of educators and how they perceive their assessment systems, a crucial factor associated with using assessment and accountability information effectively is understanding how these data are associated with improvements in student learning. The next set of analyses focuses on this factor by comparing educators who perceived their students' achievement to be static or declining with those who perceived student achievement in their schools to be improving within each of the three groups. These comparisons indicate different perceptions across educator groups regarding implementation of the effective assessment components (See Table 3).

Table 3. Comparison across Components of Educators from Improving and Static or Declining Districts/Schools

	Superintendents			Principals			Teachers		
	Declining or Static Levels of Ach. (N=16)	Improving Ach. Levels (N=32)	Effect Size	Declining or Static Levels of Ach. (N=34)	Improving Ach. Levels (N=75)	Effect Size	Declining or Static Levels of Ach. (N=35)	Improving Ach. Levels (N=58)	Effect Size
High expectations	NA	NA	NA	4.09 (0.81)	4.36 (0.57)	*0.39**	4.08 (0.61)	4.28 (0.60)	*0.33**
High-quality assessments aligned with standards	3.42 (0.59)	3.49 (0.66)	0.13	3.49 (0.97)	3.66 (0.74)	*0.20*	3.17 (0.74)	3.38 (0.71)	*0.29*
Alignment of resources, support and assistance for improvement	3.27 (0.76)	3.09 (0.74)	*-0.24*	3.19 (0.57)	3.68 (0.59)	***0.84***	3.34 (0.76)	3.59 (0.68)	*0.35**
Applying sanctions and rewards	3.62 (0.63)	4.16 (0.49)	***0.96****	3.44 (0.55)	3.53 (0.62)	0.15	3.27 (0.79)	3.59 (0.76)	*0.41**
District and School Personnel Data usage	3.36 (0.48)	3.81 (0.60)	***0.83****	3.71 (0.50)	3.96 (0.50)	**0.50**	3.25 (0.63)	3.96 (0.50)	***1.25****
Informing parents and community	3.97 (0.53)	4.13 (0.56)	0.55	3.27 (0.59)	3.49 (0.67)	*0.35**	3.00 (0.79)	3.36 (0.76)	*0.46**

* Significant to p<0.05. **Significant to p<0.01. Effect sizes in *Italics=Small,* **Bold=Medium**, ***Large=Bold Italics***[9, 10]

There were differences on almost every measure when comparing respondents' scores relative to changes in student achievement. Often, these differences became even more apparent when

moving from superintendents to principals to teachers. This might be due to the changes in proximity that each group of educators has with the students. In other words, teachers, having the closest relationships with students, might see more of an impact from changes in their practices relative to student achievement. Teachers are also privy to a myriad of data (e.g. course grades, homework, class participation) that can help inform their work.

The most significant differences between the different educator groups were found in how data were used to evaluate and identify strengths and weaknesses in decision making at the school and classroom levels. Medium and large effect size differences were apparent in the perceived use of assessment data to inform policy and practice between schools that were improving in student achievement and those believed to be either stagnant or declining. Teachers reported the greatest perceived differences, with superintendents also reporting a large effect size. Also noteworthy are the apparent consistencies in responses among superintendents, principals, and teachers at improving schools in terms of the degree to which they rated their level of data use. This consistency may indicate more alignment between these two groups on how data are being used within the school. Overall, these findings indicate a strong association between using data and improving academic achievement.

Of further significance are the differences in beliefs about the distribution of resources and levels of support between the academically improving and static or declining schools. It seems reasonable that if educators are using data in the ways described earlier, they might be better positioned to make informed decisions about allocating and aligning resources to improve student achievement. Thus, the large and medium effect size differences between the principal and teacher groups, respectively, was not surprising. Nor is it surprising that principals and teachers in improving schools reported using data to inform their decisions and to support teachers in using data in the classroom. A negative effect size comparing the superintendent groups, however, was unexpected. These administrators in the static and declining group communicated a significantly higher belief that they have the necessary support needed to improve student performance than those in the improving group. Differences in the kinds of support expected to improve student achievement between superintendents representing the two groups may help to explain this unanticipated finding. Superintendents who are implementing data-driven goals in their schools and who are examining the needs of all their schools to define the types of supports (specialized trainings, supplementary education programs, etc.) may see a greater disparity in resources between what they need to meet those goals and their current level of resources than their colleagues whose expected support systems are not as dependent on student performance (Ferguson, 2002; Ogbu, 2003). Though superintendents were not surveyed on this component, teacher and principal differences with respect to clearly communicated school expectations that all students perform at high standards indicate that educators of improving schools have significantly stronger beliefs that their students can reach high levels of achievement. Also, teachers and principals in improving schools were fairly consistent in their beliefs that high expectations were a part of their school culture. This finding is especially noteworthy because it is a fundamental component of the standards-based reform movement.

The application of rewards and sanctions showed varying degrees of differences between administrators and teachers. Superintendents in districts that were perceived to be improving were significantly (d=0.96) more likely to use data to evaluate school progress and to attach consequences to those results than were superintendents of static or declining schools. Teachers also showed significant differences in the belief that assessment data was used to measure school effectiveness and sanctions were imposed on those schools not meeting standards. Principals, on the other hand, indicated no significant differences in their beliefs about the implementation of rewards and sanctions in their schools. This pattern is interesting because the results of the first set of analyses showed that teachers overall rated this component significantly lower than

principals. However, when looking at the data disaggregated by the changes in achievement, the teachers and principals in improving schools rated the use of sanctions and rewards very similar to each other. This may indicate a degree of consistency in 'improving' schools — teachers and administrators view the application of rewards and sanctions to the same degree.

Educators of improving schools and districts also reported supplying more information to parents and community members about their school's progress. Superintendents showed dramatic differences and a medium effect size regarding the belief that policies for communicating results to the public were established and implemented. Principals and teachers also showed significant differences between the two groups with small effect sizes. However, the trend noted in the first set of analyses was still apparent— overall, superintendents rated this component higher than principals and teachers even when consideration is given for changes in student achievement. This might indicate the need for an increased focus on this factor at the school level, as indicated above.

Analysis of the Nested Data

The main purpose of these analyses was to provide additional evidence to support the initial set of findings from the overall sample. Analyses were conducted to examine the degree to which respondents viewed the components of effective systems consistently within the same educational system. In other words, did teachers view the use and quality of the components in a ways that were similar to superintendents' and principals' views? Although this analysis reduced the sample sizes for superintendents and principals, an analysis of the nested data provided more robust evidence to examine the responses being similar or different across the groups.

Nested Sample. The first step in this analysis was to select a sub-sample of nested data from the larger dataset. This sample selection would allow us to examine the data from a system perspective. In other words, we only looked at responses for which we had a complete set of school and district responses (i.e. we had responses from teachers, their principal, and their superintendent). The entire teacher sample of 153 was used in connection with that of their principals and superintendents. This resulted in a smaller sample of administrators as compared to the original datasets, but allowed us to examine only those responses for which we had data from all three groups. Because we were working from a smaller sample of principals and superintendents, demographic data were examined to see if there were differences compared to the original populations. In terms of location, 58 percent of superintendents and 48 percent of principals reported that their districts or schools were rural compared to the entire sample. This is not substantially different from percentages reported in the original sample or percentages of rural schools and districts in the surveyed states. With regard to SES, administrators reported, on average, that their school or district had 40 percent of their students qualify for free or reduced-price lunch. Although slightly higher than the overall averages for their states, this is similar to the responses of administrators in the overall sample. Percentages of minority students were also similar to the general population and responding principals represented a range of school types (elementary, middle, and high). These data led us to conclude that there seemed to be no apparent sampling biases due to the smaller sample of administrators in the nested sample.

Results. The results presented in Table 4 are consistent with results of the complete sample presented in Table 2. One might expect to see more consistency between the three groups of respondents when consideration is given to the fact that the respondents come from the same districts and schools. In fact, some of the measures show marked differences between the respondents. For example, there are bigger differences between principal and teacher perceptions on the degree to which 'High Expectations' were present in their school. A small effect size

(0.37) indicates that principals in the nested sample were more likely than teachers to feel that 'High Expectations' were part of their school's culture. When considering the entire sample, there were no differences between teachers and principals' perceptions.

Table 4. Comparisons across Effective Characteristics for the Nested Sample

Variable	Superintendent Mean (SD) N=19	Principal Mean (SD) N=27	Effect Size (Superintendent - Principal)	Teacher Mean (SD) N=153	Effect Size (Principals – Teachers)
High expectations	NA	4.39 (0.50)	NA	4.19 (0.59)	*0.37*
High-quality assessments	3.42 (0.63)	3.79 (0.72)	**-0.55**	3.29 (0.73)	**0.69****
Alignment of resources, support, and assistance for improvement	3.25 (0.69)	3.51 (0.57)	*-0.41*	3.48 (0.73)	0.05
Applying sanctions and rewards	3.95 (0.54)	3.48 (0.56)	**0.85****	3.22 (0.79)	*0.39**
District and school personnel data usage	3.72 (0.42)	3.91 (0.56)	*-0.39*	3.44 (0.63)	**0.79***
Informing parents and community	4.11 (0.54)	3.33 (0.77)	**1.19****	3.20 (0.80)	0.17

* Significant to p<0.05. **Significant to p<0.01. Effect sizes in *Italics=Small,* **Bold=Medium**, ***Large=Bold Italics***[9, 10]

Another important finding from the analysis of the nested sample is the lower rating between teachers and principals regarding the measure of *High Quality Assessments*. Principals tended to rate the quality of the state assessment system much more highly than did teachers. As mentioned earlier, this might be due to the fact that state assessment data is often reported at the school level making it easier and more useful for principals who need big picture information. Also, large-scale, statewide tests lack the timeliness and refined, diagnostic information that teachers need to begin to implement change.

DISCUSSION

This study provides insights into how educators perceive the quality and utility of their accountability systems as well as the degree to which they are using these systems to inform decisions in critical areas such as allocating resources and adapting their instruction. Although the descriptive nature of this research precludes the issuing of causal claims, some conclusions can be drawn based on the associations and differences between and across educator groups. Overall, the results indicate that educators view the utility and quality of their systems in a positive light. However, large disparities become apparent when comparisons are made across the administrator and teacher groups; these may be a reflection of how group members interact with the accountability system and the data it produces.

Each of the three surveyed groups have different uses for data, different levels of accountability, different responsibilities and priorities, and ultimately each group impacts the educational system in different ways. Superintendents, for example, may have more at stake *personally* when confronting issues of accountability. In particular, superintendents' job responsibilities, by their very nature, entail being the spokesperson for their district. Superintendents arguably hold the primary position of addressing the public at large about accountability. They are accountable for answering questions about how tax dollars are spent, answering to an elected school board, and

ensuring that their district meets federal requirements. And, as accountability data are increasingly being reported in the media, superintendents are forced under an even higher powered microscope; therefore, it is not surprising that superintendents rated the area of informing parents and the community of assessment and accountability results very high given their high profile responsibility for addressing the public's concerns. Also due to their public role, superintendents are perhaps the most acutely aware of impending penalties and incentives and hence responses to issues such as the degree sanctions and rewards are systematically employed reflect relatively high levels of concern and awareness. Accountability plays a large role in their day-to-day lives and they are the key players in ensuring the public that tax dollars are being wisely spent. Therefore, they are responsible for addressing issues and concerns if schools in their districts to not meet such criteria as adequate yearly progress (AYP).

Superintendents also reported lower degrees of resources and support for improvement. This was evident in the entire sample, for the nested sample, and surprisingly for superintendents in districts that were improving in student achievement. Superintendents are feeling the disparity between trying to match resources to identified needs, and the data seem to indicate that superintendents that are leading the most effective districts in terms of student achievement are feeling the pinch even more.

Because principals play a crucial role as intermediaries between district administration and teachers, their responses often served as a middle ground between the other two groups of respondents. For example, principals reported that sanctions and rewards were used and that they communicated with parents and the community. These results were significantly lower than those of superintendents but significantly higher than those of teachers. The findings seem to suggest that there may be a decreasing level of perceived responsibility as respondent roles (superintendent, principal, and teacher) move from administrative tasks to the day-to-day education of children. It's possible that these issues are more salient for the superintendent and are seen as less relevant to building level educators. However, there were some areas that did not follow this trend such as data usage and the quality of the state assessment, both of which were rated fairly high. Thus, principals might feel that assessment and accountability data adequately meet their needs at the school level.

Also noteworthy was that teachers reported following school and district policies regarding using data to inform their practices. However, further discussion of data usage showed relatively superficial implementation as compared to that of administrators. For example, when describing the tools that they used to analyze data, teachers stated that they just looked at the test results but failed to mention utilizing resources such as hardware or software to assist them in data use. This difference could be due to many factors, including lack of training, lack of time, or lack of interest in using data. It also is possible that accountability data are not seen as relevant to teachers because the information seems far removed from their day-to-day practices. Teachers noted that often they get data too late to use it to impact their instruction, or don't receive data at all from their state accountability system. These concerns coupled with a lack of resources (time, training, or personnel) to support data use might help explain how the data become irrelevant for them. Moreover, teachers have increased access to different, alternate types of data such as class participation, homework, and student/teacher interactions that ultimately might be more informative to meeting their needs. Conversely, the responses of principals and superintendents indicated a higher level of data usage, which could be due to the fact that traditional accountability data are more aligned with the responsibilities associated with their respective roles, which often entails looking at trends using data in aggregate forms.

CONCLUSIONS

What do these differences mean to schools and districts trying to raise student achievement? It seems that increased communication at all levels of the education system is important to ensure consistency in how policies are understood and implemented. Surely it cannot be assumed that teachers will implement a policy instituted at the district level without dialogue about policy intent and policy meaning, along with support in the way of resources to analyze data. Thus, when different stakeholder groups engage in a dialogue, district directives not only serve district accountability needs but also align with — or provide for — the needs of teachers in their day-to-day teaching.

The following are some issues that should be considered when evaluating assessment and accountability systems:

- Using data to make informed decisions about instruction is a crucial component to an effectively functioning accountability system. There were large differences on this measure when schools improving in student achievement were compared to those who were static or declining. Educators from schools and districts that were improving were more likely to use data to inform their work. Yet differences persist among superintendents, principals, and teachers regarding the degree to which this is occurring in schools. In particular, the data indicate that there is a disconnect between administrators and teachers and that administrators need to support teachers in ways that make data usage more effective and easier. This may include developing better supplemental sources of data in addition to the statewide test to provide teachers with diagnostic data on their current students.

- Training for teachers, principals, and administrators needs to be developed that facilitates dialogue about how to best allocate resources, identifies weaknesses in the curriculum, supports teachers and effectively monitors policy implementation.

- Educators in schools and districts that have not seen improvements in student achievement might critically examine their practices regarding using data to inform their work, communicate with parents, and allocate resources, as well as ensure that they are in fact holding to the philosophy that all students can become proficient in meeting rigorous academic standards.

An important question remains about how best to support the use of data given limited resources and competing demands for time. These issues are complex, but addressing them becomes essential as the focus on student progress and achieving yearly growth goals sharpens.

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.

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.

Englert, K., Fries, D., Goodwin, B., Martin-Glenn, M. & Michael, S. (2004). *Understanding how principals use data in a new environment of accountability.* Aurora, CO: Mid-continent Research for Education and Learning.

Englert, K., Fries, D., Martin-Glenn, M, & Michael, S. (2005). *Understanding how teachers use data in a new environment of accountability.* Aurora, CO: Mid-continent Research for Education and Learning.

Ferguson, R.F. (2002). What doesn't meet the eye:Understanding and addressing racial disparities in high- achieveing suburban schools. Cambridge, MA: Havard University, John F. Kennedy School of Government. (ED 474 390)

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.

Heibert, J, Stigler, J. W., Jacobs, J. K., Givvin, K. B., Garnier, H., Smith, M., Hollinsworth, H., Manaster, A., Wearne, D., Gallimore, R. (2005) Mathematics teaching in the United States today (and tomorrow): results from the TIMSS 1999 video study. *Educational Evaluation and Policy Analysis:* 27(4) 111-132.

Ogbu, J.U. (2003). Black American students in an affluent suburb: A study of academic disengagement. New Jerssey: Lawrence Erlbaum. (ED 476 118)

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.

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

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.