



ACCOUNTABILITY SYSTEMS: A COMPARATIVE ANALYSIS OF SUPERINTENDENT, PRINCIPAL, AND TEACHER PERCEPTIONS

KERRY ENGLERT, PH.D., DAWN FRIES, MYA MARTIN-GLENN, MA, BRUCE DOUGLAS, MS
Mid-continent Research for Education and Learning (McREL)

A key assumption of NCLB appears to be that assessment data in and of itself can foster or promote change. Specifically, the supposition is that by requiring assessment data to be reported yearly, schools will be motivated - and will have the ability - to address those areas where student achievement is lagging. This assumption rests on the notion that educator competence in understanding and utilizing such data will result in academic success. Testing this assumption with empirical evidence is an important component of researching the efficacy of current accountability policies and practices in general. Over the past three years we have been involved in a series of empirical examinations of accountability. Each of these studies has been aimed at gathering varied perspectives on and about accountability, ranging from superintendents to principals to teachers. Our research examines education accountability at three interconnected layers: district administrators, principals, and teachers. This nested data set (superintendents were surveyed, as were their principals, and their principals' teachers) allows for not only an examination of the perceptions and reflections of the members of each group but also for an evaluation of the consistency of those beliefs across the members of the educational community. This study will present findings from research projects that speak to each of these levels, focusing on how each understands education accountability and how those meanings are consistent across groups and to what degree.

Englert, K., Fries, D., Martin-Glenn, M., & Douglas, B. (2007). Accountability Systems: A Comparative Analysis of Superintendent, Principal, and Teacher Perceptions. *International Journal of Education Policy and Leadership* 2(4). Retrieved [DATE] from <http://www.ijep.org>.

Introduction

No Child Left Behind (NCLB) was enacted in the public interest as a means to improve student achievement in the United States. A key assumption of NCLB appears to be that assessment data in and of itself can foster or promote change. Specifically, the supposition is that by requiring assessment data to be reported yearly, schools will be motivated—and will have the ability—to address those areas where student achievement is lagging. This assumption rests on the notion that educator competence in understanding and using such data will result in academic success. For example, at the district level, the expectation appears to be that administrators, given a set of indicators, will have the capacity and resources to assess the situation and use this information as they establish policies and press for changes that will lead to improved student per-

formance. Testing this expectation with empirical evidence is an important component of researching the efficacy of current accountability policies and practices in general. Additionally, this analysis provides comprehensive view of these policies and practices, which is essential to understanding of the issues to implementation (Desimone, 2006, in press).

Over the past three years, we have been involved in a series of empirical examinations of accountability. Each of these studies has been aimed at gathering varied perspectives on and about accountability, ranging from superintendents to principals to teachers. Drawing from these examinations of accountability, we maintain that accountability serves many different functions. Our research examines education accountability at three interconnected layers: district administrators, principals, and teachers. This nested data set (superintendents were surveyed, as were



SIMON FRASER
UNIVERSITY



their principals and their principals' teachers) allows for not only an examination of the perceptions and reflections of the members of each group but also for an evaluation of the consistency of those beliefs across the members of the educational community. This study will present findings from research projects that speak to each of these levels, focusing on how each understands education accountability and how those meanings are consistent across groups and to what degree.

Seven Characteristics of Effective Accountability Systems

This article summarizes a comprehensive literature review of how education researchers define high-quality assessment systems (see Goodwin, Englert, & Cicchinelli, 2003). Surveys based on this literature review, as well as additional research, were developed and administered to superintendents, principals, and teachers in order to gain an understanding of how they view the quality of accountability systems and how they are using these data to impact change.

The effectiveness of schools' accountability policies is of paramount concern to the success of NCLB. If schools are not actively engaged in effectively using accountability data, generating the increases in student achievement required by this legislation becomes unattainable. Goodwin, Englert, & Cicchinelli (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

Of these 12 characteristics, the following seven were identified as being particularly relevant to the way educators use and perceive their accountability systems: (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) data that are readily understandable to the public. These elements, or a subset thereof, were selected and used as the foundation for constructing a set of three surveys given to superintendents, principals, and teachers. Grounding the research study in these characteristics allows for an examination of the overall effectiveness of current school policies and practices. By examining these characteristics, a better sense can be gained of how

and if accountability systems are being used in ways that might yield increases in student learning and 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 should meet standards. These policies called for teachers and administrators to hold students, both advantaged and disadvantaged, to the same high standards.

According to Baker, Linn, Herman, and Koretz, "Accountability systems should include the performance of all students, including subgroups that historically have been difficult to assess" (2002, p. 2). Research has found that maintaining and communicating a schoolwide academic press for achievement is a contributing factor to helping low-achieving students reach higher standards and, consequently, to creating an effective school (Creemers, 1994; Kannapel & Clements, 2005; Marzano, 2000; Teddlie & Reynolds, 2000).

High-Quality Assessments Aligned with Standards

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 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 (Baker et al., 2002; Buckendahl, Impara, & Plake, 2002; Guth et al., 1999).

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 note that "recent research is now

¹ The 12 essential characteristics identified by Goodwin, Englert, and Cicchinelli (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) data that are readily understandable to the public; (9) flexibility and fairness to allow for local differences and creativity; (10) balanced, comprehensive design; (11) stakeholder support and engagement; and (12) fairness provisions.

supporting the hypothesis that resource levels can make significant differences in achievement, and that disadvantaged students probably benefit more from increased educational resources" (1998, 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 found that states with a lower average socioeconomic status can show significant gains in student achievement "through modest increases in resources, if allocated to specific programs" (2000, p. 101).

Sanctions and Rewards Linked to Results

There are many examples of accountability models that use sanctions and rewards to spur school improvement efforts. In a report examining the gains in elementary and junior high school test scores in Texas and North Carolina, Grissmer and Flanagan 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" (1998, p. iv). Accordingly, *Education Week's* ratings of state accountability systems include whether the state "holds schools accountable for performance" (2002, p.12) through rewards and sanctions, including closure, reconstitution, transfers, and withholding of funds. Sanctions and rewards are also 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. More recently, Nichols, Glass, and Berliner (2006) found contradictory evidence in that high-stakes testing might increase student performance in some areas such as 4th and 8th grade mathematics where drill and teaching to the test might be easier, whereas reading scores were not affected by the degree of pressure felt by the accountability system.

Multiple Measures

Many accountability guidelines call for the use of multiple measures to ensure a more complete and accurate assess-

ment of students, teachers, and schools (Baker et al., 2002; Sirotnik & Kimball, 1999). Sirotnik 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" (1999, 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).

District and School Personnel Data Usage

Measuring in multiple ways student achievement and school progress toward goals is important; equally critical is strategically using these data to diagnose problems and work toward solutions. Research has found that principals want to use data to provide instructional leadership in a more informed way (Torrence, 2002). Conversely, Fuhrman (1999) found that many schools labeled as "inadequate" did not use test results diagnostically. By using data to evaluate curricula, staff, and students, principals can focus their efforts and resources in the areas that are the most deficient.

Informative to Parents and the Community

Another commonly identified element of accountability systems is their ability to inform students, parents, and teachers about student progress and status. In its evaluation of state accountability systems, *The Princeton Review*, 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" (2002, p. 6). According to Walberg, one principle that encourages the effective use of accountability system data is "user-friendliness" (2002, p. 158). Reports about student and school progress should be useful and understandable to all interested parties regardless of their level of technical knowledge.

These seven characteristics of effective accountability systems outlined here provided a theoretical framework for this study. The method derived from this framework allowed for an examination of the consistency of perceptions across the characteristics. Our hypothesis was that there would be no differences in perceptions of these characteristics across the three groups. In other words, if a characteristic in the system was deficient, all three stakeholder groups would view this inadequacy in a similar fashion. Conversely, if a characteristic was of high quality, the stakeholder groups would consistently rate it as such. The goal

of implementing an effective accountability system might only be met if important characteristics are consistently implemented and perceived across different levels in the educational system from administration to the classroom.

Research Questions

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 use policies and practices that research and literature have identified as being effective?
- Do policies and practices of principals and teachers in high-performing schools differ from those of low-performing schools?

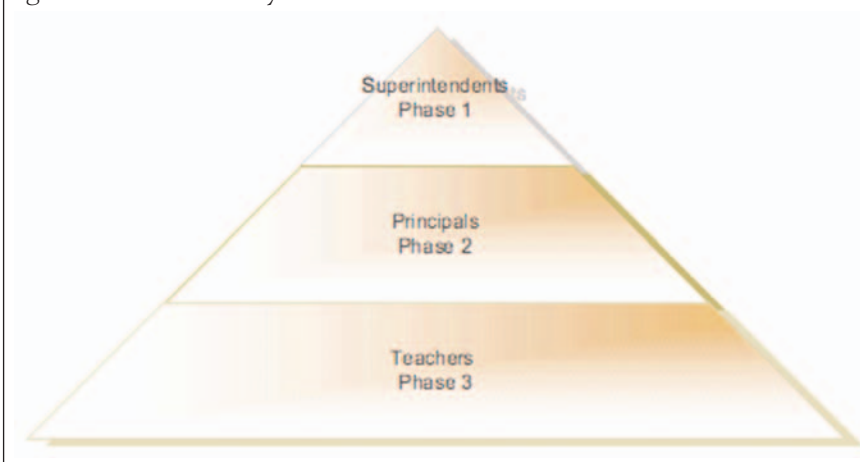
Method

A series of three surveys were constructed to measure the degree to which the practices of superintendents, principals, and teachers reflect the seven effective characteristics identified above, as well as how these educators perceive school and district level policies and other assessment related issues. Each survey measured educators' perceptions and policies on accountability through closed-ended questions and generated detailed contextual information about these elements via open-ended items. Items and response sets were kept as consistent as possible to better ensure comparability across stakeholder groups (Fowler, 1995, as cited in Desimone, 2006). Measures were developed from the closed-ended items to provide more reliable indicators of the constructs and to ensure that the practices related to each of the elements were more accurately interpreted. Prior to distributing each survey, it was pilot-tested with a small group of current and former teachers and administrators. This pilot testing resulted in clarification of the wording of existing items and inclusion of additional items. A final group of surveys was developed and reviewed by the project team based on reviewers' comments.

Sampling

As described in earlier reports (Englert, Fries, Goodwin, & Martin-Glenn, 2003; Englert, Fries, Goodwin, Martin-Glenn & Michael 2004; Englert, Fries, Martin-Glenn & Michael 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 a 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 represented 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 students and English language learners.² 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 surveys³ (see Figure 1). For some of the analyses the entire sample from each population was used. For other analyses, a sample was pulled from the larger sample that represented all of the teachers (N=153), only the principals of those teachers (N=27), and

Figure 1. Phases of Study Data



² Having the chief state school officers select the sample may have resulted in sampling bias; however, because they were most familiar with their states and districts, it was determined that they were best qualified to purposively select appropriate districts.

³ For more details on sampling issues, see individual reports (Englert et al., 2003; 2004; 2005).

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.

Analyses

Analysis of the Nested Data

Analyses were conducted to examine the degree to which respondents viewed the components of effective systems consistently within the same school district. In other words, did teachers view the use and quality of the components in a similar way relative to the superintendents and principals? Although this analysis reduced the sample sizes for superintendents and principals, an analysis of the nested data provided more robust evidence on the similarities and differences across the groups.

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 allowed us to examine only those responses for which we had data from all three groups. Because we were working with a sub-sample of principals and superintendents, demographic data were examined to see if there were differences compared to the original populations. In terms of location, percentages of minority students, and percentages of students receiving free or reduced price lunch, there were not substantially different percentages from those reported in the larger samples or from the percentages in the population. These data led us to conclude that there seemed to be no apparent sampling biases from the smaller sample of administrators in the nested sample.

When comparing groups, we first compared principal and superintendent responses and then compared principal and teacher responses. The rationale for only examining differences between these groups was that we believed superintendents and teachers had roles that were too far removed from each other and that studying the differences through the role of the principal would provide an adequate bridge. This would also limit the number of t-tests that were performed, in turn limiting the chance of committing a Type I error.

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, teachers rated many of the different components of effective accountability systems the lowest of the three groups (see Table 1, page 6).

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 work together to make the necessary improvements to address and correct any weaknesses in the education of all students. This goal can 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. The results indicate that overall, teachers rated their usage of data much lower than did principals and superintendents. Mean differences between teachers and principals, for example, resulted in statistically significant and medium effect size (0.79) differences. In comparison, mean differences between principals and superintendents suggested nonsignificant and small effect size (-0.39) differences.

Respondent comments suggest several reasons why teachers rated their level of data usage 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 commented that “[there are] too many demands for time, too much pressure on students and staff.” 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. . . .

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 (.050)	NA	4.19 (.059)	0.37
High-quality assessments	342 (.063)	379 (.072)	-0.55	329 (.073)	0.69**
Alignment of resources, support, and assistance for improvement	325 (.069)	351 (.057)	<i>-0.41</i>	348 (.073)	0.05
Applying sanctions and rewards	395 (.054)	348 (.056)	0.85**	322 (.079)	0.39*
District and school personnel data usage	372 (.042)	391 (.056)	<i>-0.39</i>	344 (.063)	0.79*
Informing parents and community	411 (.054)	333 (.077)	1.19**	320 (.080)	0.17

* Significant to $p < .005$. **Significant to $p < .001$. Effect sizes in *Italics*=Small, **Bold**=Medium, **Large**=**Bold Italics**^{5, 6}

Principals also noted that their teachers need more training on data use and data-driven instructional strategies. According to one principal, the top three needs that school has around data use are, “Teacher training to interpret data. Teacher training in the application to teaching. Teacher training in how to better use classroom formative and summative data for instruction.” 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 believe 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, “I don’t want to use it [the state assessment]. I would like to focus on my class in my community without expectation from NCLB.” This lack of confidence in current accountability systems may also contribute to teachers working 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, their most frequent response was simply citing 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 becomes cumbersome, if not impossible.

In another question, respondents were asked to describe the policy or expectation in their school that

⁴ For all items, a 5-point Likert scale was used, in which 1 indicates the negative end of the scale and 5 indicates the positive end (i.e., to no extent versus to a great extent or strongly disagree to strongly agree).

⁵ 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.

⁶ Effect sizes are generally considered to be “small” if they are between 0.20 and 0.50, “medium” if they are between 0.50 and 0.80, and “large” if they are greater than 0.80 (Cohen, 1988).

teachers use assessment data to inform their classroom practices. Overwhelmingly, the administrators indicated that teachers are expected to plan their instruction and curriculum for their students based on the needs reflected in the data. One superintendent stated that teachers in that district “are expected to meet in grade levels and content areas to analyze the data, and incorporate their findings into instruction.” 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 school’s progress. Although superintendents had relatively high ratings on the extent to which parents were informed of assessment results, principals’ mean rating was significantly lower, and the rating of 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. “Individual student assessments are made available to parents” and “policy by handbook” were typical comments in many responses.

Principals and teachers tended to be more specific about how student results are shared with parents. The most often cited method was direct teacher–parent communication at conferences and meetings. Though responses indicated 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 often cursory (e.g., “Teachers are to share assessments results with parents,” “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, and only a few respondents mentioned sharing strategies with parents on how they can best support their child’s learning at home. Moreover, the above findings 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. Superintendents were more likely to be using data to monitor the progress of schools towards data driven goals and to attach consequences to those results compared to principals. 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 the unprecedented challenges of all students being grade-level proficient in reading or language arts and mathematics by 2014, as mandated in NCLB, are placing demands on superintendents to respond to issues regarding sanctions but that these demands are not yet critically felt at the school 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 because often it is assumed that teachers working in the trenches might feel the most pressure from not having enough supplies, computers, curricular materials, and staff. These results, however, could 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.

Finally, large differences were detected between principals’ and teachers’ perceptions of the degree to which *High Expectations* are embodied by staff at their schools. Though the magnitude of the effect size difference was small (0.37), findings indicate that principals in the nested sample were more likely to feel that *High Expectations* were part of their school’s culture relative to teachers. This might be due to the different roles principals and teachers play in the education of students. Whereas teachers experience direct interaction with most of their students on a daily basis, principals experience this interaction with only a few.

Analysis of Responses from Educators Based on Improvements in Student Achievement

While these findings provide a picture of the overall consistency between groups of educators and how they perceive their assessment systems, a crucial factor associated with utilizing assessment and accountability information effectively is understanding how these data are associ-

ated 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. For these analyses, *all* district and school administrator responses were used, regardless of whether respons-

Table 2. Comparison of Educators from Districts/Schools with Improving and Static or Declining Student Achievement Levels⁷

	Superintendents			Principals			Teachers		
Variable	Declining or Static Levels of Achievement 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
	Mean (SD)	Mean (SD)		Mean (SD)	Mean (SD)		Mean (SD)	Mean (SD)	
High expectations	NA	NA	NA	4.09 (.081)	4.36 (.057)	0.39*	4.08 (.061)	4.28 (.060)	0.33*
High-quality assessments aligned with standards	3.42 (.059)	3.49 (.066)	0.13	3.49 (.097)	3.66 (.074)	0.20	3.17 (.074)	3.38 (.071)	0.29
Alignment of resources, support, and assistance for improvement	3.27 (.076)	3.09 (.074)	-0.24	3.19 (.057)	3.68 (.059)	0.84**	3.34 (.076)	3.59 (.068)	0.35*
Applying sanctions and rewards	3.62 (.063)	4.16 (.049)	0.96*	3.44 (.055)	3.53 (.062)	0.15	3.27 (.079)	3.59 (.076)	0.41*
District and school personnel data usage	3.36 (.048)	3.81 (.060)	0.83*	3.71 (.050)	3.96 (.050)	0.50**	3.25 (.063)	3.96 (.050)	1.25*
Informing parents and community	3.97 (.053)	4.13 (.056)	0.55	3.27 (.059)	3.49 (.067)	0.35*	3.00 (.079)	3.36 (.076)	0.46*
* Significant to p<0.05. **Significant to p<0.01. Effect sizes in <i>Italics</i> =Small, Bold =Medium, <i>Large</i> = Bold Italics ^{8,9}									

⁷ For all items, a 5-point Likert scale was used, where 1 indicates the negative end of the scale and 5 indicates the positive end (i.e., to no extent versus to a great extent or strongly disagree to strongly agree).

⁸ 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.

⁹ Effect sizes are generally considered to be "small" if they are between 0.20 and 0.50, "medium" if they are between 0.50 and 0.80, and "large" if they are greater than 0.80 (Cohen, 1988).

es were received from other educators in their school or district. These educator comparisons indicate different perceptions regarding implementation of the effective assessment components across the educator groups (see Table 2).

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 and most direct 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 non-state assessment 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 difference (1.25), with superintendents also reporting a large effect size (0.83). Also noteworthy are the apparent consistencies across superintendents, principals, and teachers at improving schools in terms of the degree to which they rated their data usage. 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 the 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 significant effect size differences between the principal (0.84) and teacher (0.35) groups were not surprising. Nor is it shocking that principals and teachers from 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 (-0.24), however, was unexpected. These administrators in the static and declining group communicated a slightly 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 static/declining group and those from the improving group may help to explain this unanticipated finding. Superintendents who are using data-driven goals and who are examining the needs of all their schools to define the types of supports (specialized trainings, supplementary education programs, and so on) 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 in improving schools have significantly stronger beliefs that their students can reach high levels of achievement than educators in static or declining schools. 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 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 at improving schools and districts also reported supplying more information to parents and community members about their schools' progress.

Superintendents showed dramatic differences and a medium effect size regarding the belief that policies for communicating results to the public (0.55) were established and implemented. Principals and teachers also showed significant differences between the two groups with small effect sizes (0.35 and 0.46, respectively). However, the trend was still apparent that was noted in the first set of analyses: 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 previously.

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 differently with the accountability system and the data it produces. These findings show even greater differences than past research conducted on policy views of school and district stakeholders (Desimone, 2006) who found that stakeholders' perceptions are different in areas such as consequences of policies but similar in perceptions around barriers to implementation.

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. These administrators arguably hold the primary position of addressing the public at large about accountability. Additionally, superintendents 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. This public role may impose on superintendents the need to be acutely aware of impending penalties and incentives as well as the consequences they may bring. Thus, responses to issues such as the degree to which sanctions and rewards are systematically employed may reflect relatively high levels of concern and awareness among this type of administrator. It is evident accountability plays a large role in the day-to-day lives of superintendents, and they are the key players in assuring the public that tax dollars are being wisely spent. Therefore superintendents are responsible for addressing issues and concerns if schools in their districts do not meet such criteria as adequate yearly progress.

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. Results suggest superintendents are feeling the disparity between trying to match resources to identified needs. The data further seem to indicate that superintendents representing the most effective districts in terms of student achievement are feeling the pinch even more.

Because principals hold a crucial role as intermediary between district administrators and teachers, their responses often reflected a middle ground between the responding superintendent and teacher groups. For example, principals reported a significantly lower utilization of sanctions and rewards than did superintendents but a significantly higher usage rate than was reported by teachers. Similar results were found for the measure relating to communication with parents and the community. These findings seem to suggest there may be a decreasing level of perceived responsibility as respondent roles move from administrative to one more involved with 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, some areas 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 compared to that of administrators. For example, when describing the

tools they used to analyze data, teachers stated that they just looked at the test results but failed to mention using 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 cited that often they get data too late 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 usage, 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 and 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.

Recommendations

The results from the study show inconsistencies across different groups of educators in how assessment and accountability systems are effectively implemented. The following are recommendations that can assist educators with addressing the issues discovered and revealed in this study. 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. Most importantly, additional research is needed to address practical issues regarding implementation of effective accountability systems across stakeholder groups.

- Using data to make informed decisions about instruction is a crucial component to an effectively functioning accountability system. Educators from schools and districts must use data more to inform their work, supporting 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.
- Teachers need to be provided the resources to analyze data so the process is efficient and aligned with their needs.

- Training for teachers, principals, and administrators needs to be developed that allows for a facilitation of a dialogue about how to best allocate resources, to identify weak areas in the curriculum, to support teachers, and to effectively monitor policy implementation.
- Educators in schools and districts that have not seen improvements in student achievement might critically examine their practices around using data to inform their work, to communicate with parents, to allocate resources, as well as to ensure they are in fact holding to the philosophy that all students can become proficient in rigorous academic standards.
- 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.
- Increased communication at all levels of education is important to ensure consistency in how policies are understood and implemented. A policy that is instituted at the district level cannot be assumed to be used by teachers unless there is dialogue about the intent and the meaning.

An important question remains about how to best support the use of data with limited resources and competing demands for time. These issues are complex but addressing them is essential to ensuring student progress and meeting yearly growth goals.

Authors' Note

This article has been funded 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.

References

- 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.
- Desimone, L. M. (2006). Consider the source: Response differences among teachers, principals, and districts on survey questions about their educational policy environment. *Educational Policy*, 20(4), 640-76.
- Education Week. (2002). Quality Counts 2002: Building blocks for success [special report]. *Education Week*, 21(16), 18-19.
- 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, October). *Understanding how principals use data in a new environment of accountability* (REL Deliverable #2004-09). 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-achieving suburban schools*. Cambridge, MA: Harvard University, John F. Kennedy School of Government. (ERIC Document Reproduction Service No. ED474390)
- 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* (rev. ed.) (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., & Calvo, N. (1999). *Evaluation of California's standards-based accountability system*. Menlo Park, CA: WestEd and Management Analysis and Planning.
- Kannapel, P. J., & Clements, S. K. (2005). *Inside the black box of high-performing high-poverty schools*. Lexington, KY: Prichard Committee for Academic Excellence.
- 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.
- Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives*, 14(1). Retrieved November 6, 2006 from <http://epaa.asu.edu/epaa/v14n1>
- Ogbu, J. U. (2003). *Black American students in an affluent suburb: A study of academic disengagement*. New Jersey: Lawrence Erlbaum. (ERIC Document Reproduction Service No. ED476118)
- The Princeton Review. (2002). *Testing the testers 2002: An annual ranking of state accountability systems*. New York: Author.
- 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). *Principal use of data: A national perspective*. Doctoral dissertation, Virginia Polytechnic Institute. Retrieved November 6, 2006, 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.