School-Based Accountability Systems: Data Sources for Teacher Development and School Improvement

Maureen M. Marshall

Abstract

First, this paper identifies a number of public policy issues and events that spurred policymakers to initiate broad-based educational reform aimed at holding schools more accountable. Second, it identifies and describes the relationships between and among internal accountability, external appraisal, and new legislative requirements to use "hard" data derived from external and school-based accountability systems to inform decision-making. Third, it attempts to explain the ethical and legal issues related to the ever-expanding efforts to hold schools and those who work and learn in them more accountable for results. Finally, it offers possible sources for further investigation and analysis that might assist the reader in gaining a greater understanding of the issues highlighted here.

Introduction

The impetus for writing this paper comes from three developments in the field of education: (a) the ever-deepening public policy concerns about the quality of public schools, (b) a growing interest among those outside the education community in creating more elaborate "systems" to appraise school quality and student progress, and (c) the burgeoning concern within both the legal and school communities regarding the ethical and legal implications of developing accountability and decision-making systems and attaching high-stakes consequences to the results.

Maureen M. Marshall is Superintendent of Schools for the Quabbin Regional School District in Barre, Massachusetts.
During the past two decades, the most significant and controversial development in the field of education has been the use of external assessment systems—designed by persons or agencies outside the schools—to appraise the effectiveness of public schools (Shepard, 2002). The effort to measure the outcomes of schooling using standardized achievement tests and increasingly to attach high-stakes consequences to the results has been the major focus of educational reform initiatives. Using tests in this way is largely a response to the public policy concern that, due to the poor academic performance of American students, our nation teeters on the brink of losing its competitive edge in the global economy. This concern was clearly highlighted in the Nation At Risk Report released in 1983, when the Commission wrote, “If an unfriendly power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war” (National Commission on Educational Excellence [Nation At Risk], 1983).

Since the report’s release, demands for greater accountability in K–12 public education have permeated, if not dominated, much of the conversation about public education at both the national and state levels. This public concern has provided the momentum for policymakers to promulgate far-reaching national and state-level legislation aimed at improving the quality of public schools.

For example, in response to the dire predictions in the Nation At Risk Report, many state legislatures enacted educational reform initiatives that were specifically directed at improving the quality of public schools by ensuring that public funds were more equitably distributed among all schools, by developing and implementing more rigorous curriculum standards, and by imposing state-level assessment programs designed to measure school, teacher, and student success in meeting the demands of these reform initiatives. However, these funding and educational reforms have vigorous critics, and the legislative effort to improve schools by imposing state-level accountability strategies remains an issue that threatens to unravel the tentative and fragile relationship between those who legislate and those who teach.
The intensity of discussion and debate about how best to determine the quality of schools and about who is accountable for what has increased dramatically over time. Stakeholders are sharply divided over the issues of high-stakes testing, sanctions for failing schools, performance standards, and the appropriateness of externally-imposed (state) accountability systems. The term "accountability" means different things to different stakeholders. Teachers often see accountability as working hard to meet the needs of students. Politicians see accountability as being synonymous with a student’s performance on a test. Others in the field of education advise against envisioning educational accountability solely as a system of rewards and punishment (Elmore and Rothman, 1999) and stress instead the need to differentiate between the “external appraisal” systems so popular today and “internal (school-based) accountability” systems that are designed not only to create and collect data about teaching and learning, but also to encourage the use of these data to inform professional practice and decision-making. The inherent political and ideological struggle between these two forms of accountability—”external appraisal” versus “school-based assessment for learning”—is destined to climb to new levels of ferocity due to recent mandates included in the reauthorization of the Elementary and Secondary Education Act of 2001, also known as The No Child Left Behind Act (NCLB). This act clearly favors the expansion of more rigorous external appraisal systems over the development of internal (school-based) accountability models that collect data to help practitioners improve the teaching and learning in their schools (Stiggins, 2002). In addition, the far-reaching Act mandates the extensive use of scientifically-based research to guide and inform the decision-making practices of educators.

NCLB, which passed with strong bipartisan support, requires the implementation of new practices of external appraisal and decision-making at the school level based on “hard” data (scientifically-based research). This requirement for hard data means that schools must find appropriate ways to collect meaningful data not only about the effectiveness of school practices and programs, but also about student performance. This requirement will most certainly encourage schools...
to create school-based accountability systems in order to gather this data. Developing these systems will require educators to consider what research deems to be best practice as well as the ethical and legal implications of using data to inform decisions that will have a significant impact on an individual student’s educational future.

While NCLB includes much that can assist schools in their quest to serve all students well, early evidence from states indicates that its over-emphasis on high-stakes, test-based accountability is undermining many good instructional practices and fostering numerous ineffective ones (Lewis, 2002). For example, opponents argue that standardized testing compromises the classroom environment because educators devote so much class time to "teaching to the test" (Powell, 2001).

Currently, NCLB is the focus of much debate at the school level, where practitioners struggle to understand and implement provisions that seem unrealistic and muddied by the lack of specific regulation and common sense. Some have noted that this struggle is due to the fact that each reform initiative is closely aligned with an external appraisal system that threatens educators and students with sanctions that may be life altering. Many prominent educators oppose designing an educational assessment system in which the primary motivating force is fear of failure or sanctions. Rather, they believe that the real promise lies in the development of an educational assessment system designed to improve, not just audit, student performance (Kohn, 2001; Wiggins, 1998). Richard Stiggins (2002) states it more precisely: "If we wish to maximize student achievement in the United States, we must pay far greater attention to the improvement of classroom assessment. Both assessment of learning and assessments for learning are essential. But one is currently in place, and the other is not" (Stiggins, 2002, p. 758). Others, like Kurt Landgraf of the Educational Testing Service, see the assessment of learning as a higher priority. In his testimony supporting President George W. Bush’s high-stakes testing proposal before Congress, he commented, "Results from these tests will provide important information that the American people and policymakers need to move this matter forward and to ensure significant educational
reform” (Landgraf, 2001, p. 4). Unfortunately, Landgraf makes no reference as to how such information might be helpful to either educators or students.

While it is important to distinguish among the various points of view regarding the use of accountability systems, it is also important for policymakers and educators to understand the ethical and legal context in which such systems actually exist. Those at the legislative and school levels who must develop and implement these accountability systems must also ensure that these programs honor and preserve the rights of all who are impacted by what may be well-meaning initiatives, but which also have the potential to do great harm.

Much has been said and written about the need to hold educators and students accountable. Much, too, has been said about the need to use the “hard” data gleaned from accountability systems to inform decisions about teaching, learning, and school programs. The challenge is obvious: School practitioners must develop school-based accountability systems from which to acquire meaningful data to use in improving school outcomes. Simultaneously, they must also remain ever-mindful of the ethical and legal context in which these programs must ultimately operate.

Many people think that the ethical and legal rights of students are important to consider only in the case of high-stakes testing for graduation and not to the entire K–12 academic program. However, this is a myopic view of students’ rights. School-level accountability systems will ultimately provide the data that will be used to make potentially far-reaching academic decisions about promotion, retention, grouping, and admittance into future programs of study. Teachers and administrators must therefore understand and value the ethical and legal issues related to the development and implementation of school-based accountability systems.

For example, the way in which educational assessment data are collected and manipulated must be subjected to close scrutiny in order to ensure that it is done carefully and honestly. If either process is
deficient, but the data are still used as the basis for making academic decisions, then the students affected might become the victims rather than the beneficiaries of those academic decisions.

Local school-level accountability programs must be subjected to the same level of ethical and legal scrutiny as high-stakes testing programs imposed by legislation. If not, students’ rights and educational opportunity may indeed be quietly but irrevocably sacrificed.

Questions of Significance
The following sections of this paper address these four themes:

- data-based decision-making in schools,
- the significance of school-based accountability systems to school improvement efforts,
- sources of information, and
- the ethical, legal, and due process issues that should be considered in the development and/or implementation of such accountability systems.

Why Data-Based Decision-Making
Currently, educators in the public schools are being called upon to initiate well-defined system models that rely on "hard" data to inform decision-making. Specifically, the need to develop and implement such internal accountability systems is clearly advanced in the mantra "scientifically-based research" included in NCLB. This phrase actually appears more than two hundred times within the text of the Act. Meeting this legislative mandate will require schools not only to use a more scientifically-based (data-driven) protocol, but also to redefine or develop school-level accountability systems to provide a credible data source at the school level from which practitioners can access information. Unfortunately, any existing school-level accountability systems are not as visible as those at the state level because the energy of more than a decade of educational reform has been almost solely invested in state-level, high-stakes testing programs that have become the sole source of data for school change initiatives.
Educational leaders, then, must not only ensure that school-level accountability systems are operational and reliable, but also that teachers and middle managers are skilled in gathering, analyzing, and using data to inform decisions about teaching and learning (Creighton, 2000). Accomplishing the former may well prove easier than the latter since there are quality accountability systems from which to garner insight. However, significantly changing the process by which teachers approach making professional decisions will require new knowledge, skill, and opportunities for practice.

**Importance of Data-Based Decision-Making**

Since the reauthorization of NCLB, the phrase "data-based decisions" has become a common subject of discussion among politicians and public school administrators. NCLB places great emphasis on the nature of educational research by requiring federal funding decisions to be determined by whether school programs and practices are based on "scientific research." "Every program and initiative bears this requirement, presumably as an antidote to thirty years of considerable federal investment in improving education that critics contend has yielded a succession of fads rather than proven successes" (Lewis; 2003, p. 339). Clearly, there are many in the educational community who see the focus on "scientifically-based research," as it is described in NCLB, as more politically motivated than scientifically based (Lewis, 2003).

However, it is important to recognize that many in the educational community agree that schools do benefit by using data to inform decisions about teaching and learning. In fact, many prominent educators and researchers embraced this idea before the reauthorization of NCLB (Protheroe, 2001). Notably, in 1994, Glaser and Silver described an educational future that would include the use of data to inform professional practice. They discussed an educational system in which testing was less about the sorting and selecting of students and more about offering information (Glaser and Silver, 1994).

Using data about student performance is one of the key indicators of school success identified by educational researchers (Cawelti and Protheroe, 2001). In a number of salient studies of successful schools,
researchers noted that practitioners in these schools used data about student performance to determine where schools were succeeding and where they needed to improve (Cawelti and Protheroe, 2001; Fox, 2000).

Research has also demonstrated that teachers who use data in decision-making engage with greater frequency in collegial discussions about professional practice. The use of data serves as a catalyst for more meaningful and informed discussions and collaboration (Creighton, 2001; Meyers and Rust, 2000). Individuals like Theodore Creighton have chided current practitioners for failing to use data in order to make sound instructional decisions. In his article, “Data Analysis in Administrators’ Hands: An Oxymoron?” he warns, “for too long, many school leaders have made decisions about instructional leadership with intuition and by ‘shooting from the hip.’ All too often, the decision-making process fails to include data collection and data analysis” (Creighton, 2001).

In summary, the old tools of educational decision-making—personal experience, intuition, and educational philosophy—do not seem to be enough. The passage of NCLB requires that states, schools, and practitioners accept the responsibility for accurately and promptly collecting, analyzing, and reporting data on student demographics and achievement. One might want to resist the intrusion of the federal government into a school arena already crowded with the demands and suggestions of high-profile business types, politicians, and educators. However, the idea that schools should use data to make decisions about the educational future of the children in our nation’s classrooms may actually have considerable merit (Creighton, 2000; Bernhardt, 1998).

**Accountability Systems: Sources of Important Information**

During the past two decades, many states have passed sweeping educational reform legislation aimed at improving the public schools. One of the most far-reaching and controversial aspects of these state-level reform initiatives is the inclusion of school and student accountability mandates. Generally, these accountability systems are imposed from sources outside the schools (usually a state-level bureaucracy) and are
based completely on high-stakes testing programs coupled with sanctions and rewards designed to push students and schools to perform at higher levels. Much of the policy-making around such accountability systems is dominated by this model (Scheurich, Skrla, and Johnson, 2000). Specific sanctions for students who fail to meet state-level performance standards often include grade retention or the denial of high school diplomas. If large numbers of students within a school consistently fail to meet expected levels of achievement, the school is subject to consequences, including public embarrassment as the school’s poor performance is compared to that of other schools, the removal and replacement of administrators and/or teachers, and even the take-over of the school by a state-appointed agency.

This focus on accountability has attracted the attention of members of the education community and has forced schools to adopt strategies to help students meet the state performance standards. The rush to meet these standards has left school communities believing that accountability is an external dictator to whom they must pay tribute. As a result, school practitioners have scurried to implement stopgap remedial measures in an effort to improve student performance on a single state-mandated, high-stakes test that too often becomes the sole component of the school’s accountability system.

This view of accountability does little to serve the public schools and leaves practitioners without access to a comprehensive, valid, and ultimately useful local data source. Few appreciate the important notion that internal accountability at the school level must precede external appraisal (Elmore, Ablemann, and Fuhrman, 1996). If schools are to better serve their students and meet the demands of the new federal legislation that requires data-based decision-making in the schools, an internal accountability system must be developed and implemented at the school level to provide practitioners with reliable and useful data about all school programs (Stiggins, 2002; Davie and Silva, 1999). These internal accountability systems must be available to supplement the important but limited data from the external, politically-driven accountability programs that currently hold center stage in educational accountability discussions (Elmore and Fuhrman, 2001).
Even though the standards and assessments that drive accountability are typically developed at the state level, school leaders must understand the need to drive school improvement at the local school level. School improvement can be realized by monitoring vital indicators of student and school improvement, by aligning professional development with the needs identified by analysis of the data, and by using assessment data and educational research to inform professional practice. These are the rudiments of effective school-level accountability systems that hold the most promise (Sirotnik, 2002).

**Ethical Considerations**

Many educators use the terms “ethics” and “professionalism” interchangeably. For the purposes of this discussion, the two terms will be considered synonymous. Ethical codes are published by the National Education Association (NEA) and the American Association of School Administrators (AASA) to inform the behavior of all public school teachers and administrators. These codes identify both the well-being of students as the fundamental value of all decision-making, and the need to support and protect the principle of due process and the civil and human rights of all individuals (NEA, Code of Ethics of the Teaching Profession, 2002; AASA, Statement of Ethics, 2002). However, identifying commonly understood and accepted ethical guidelines within the education profession is an elusive and difficult task. When asked, teachers and administrators have no detailed ethical code that they can recite. In fact, the literature indicates that they often have no specific knowledge of the ethical codes promulgated by their professional groups or associations. Instead, they generally note that they should do what is fair and in the best interest of their students (Shapiro and Stefkovich, 2001, p. 21). Unfortunately, what one teacher might consider fair in a matter of school discipline might not be considered fair by another. Such situations are ever present within schools and are exacerbated by the complexities of the problems that confront teachers and administrators on a regular basis.

Consider, for example, the ethical dilemma that teachers and administrators might face if they were directed to implement an accountability system that relied on a single exit exam as the sole determinant for
grade-level promotion. Would the teachers be able to discuss this situation using the ethical codes of the teaching profession as a guide? Few individuals prominent in the field of educational ethics feel that these codes are "anything other than platitudinous and perfunctory. Moreover, there is little evidence that (these codes are) taught or that most educators are aware of (their) existence" (Strike and Ternasky, 1993, p. 2). Others also note that it is doubtful that such codes have any active role in deliberations about the professional conduct of teachers and administrators in schools (Begley and Johansson, 1998). Yet dilemmas such as this are common in schools and it is important for educators to have a better understanding of the ethical guidelines of the professional organizations with which they are affiliated.

School practitioners must understand that they are the "actors" who must judiciously distribute power in their classrooms and throughout the school. They must accept this responsibility as they develop and implement school-based accountability systems designed to improve teaching and learning. They must not be blind to the ethics of their profession. There is a clear need for more considered reflection about the ethical concerns raised by controversial high-stakes testing programs and the uncertain mandates included in NCLB. Educators have little formal experience in working through their own professional ethical codes, in trying out what they might have discovered in the process, or in reflecting on solutions they reached when they were called upon to analyze various ethical dilemmas in their schools (Shapiro and Stefkovich, 2001). The new mandates of federal reforms will require educators to wrestle with ethical dilemmas and decisions about the educational future of individual students, which should not be contemplated in an ethical vacuum. Educational leaders must accept and implement "... procedures for making decisions that respect the equal sovereignty of the people" (Strike, 1991, p. 415).

**Legal Considerations**

The legal issues related to accountability systems that rely on the results of high-stakes tests to determine the educational future of students and the quality of schools have captured considerable interest and generated much analysis within the legal community. For example, if
data from an internally developed (school-based) accountability system are to be used as the primary basis for decisions about promotion, retention, grouping and determination of entry into accelerated school programs, then the validity of the accountability program, as well as how data from the program are manipulated and reported, must be subject to a level of legal and ethical scrutiny similar if not equal to that demanded for high-stakes testing programs.

This section explores the federal constitutional and statutory bases for challenges to student accountability systems. This information will highlight various legal interpretations and identify precedents regarding past, present, and potential challenges to school accountability programs.

There have been a number of legal challenges to various high-stakes accountability programs and practices. Many of these challenges have focused on claims alleging:

- a failure of the state to provide sufficient due process,
- a violation of the Equal Protection Clause of the Fourteenth Amendment,
- "state" efforts to burden a fundamental right,
- race-based classification (disparate impact claims),
- violations of Title VI (discriminatory practices), and/or
- the invalidity of educational measures or practices.

It should be noted that members of the legal community do not view each of these claims or challenges as equally significant or viable; in fact, there is considerable disagreement as to which challenges might ultimately produce the best result.

**Due Process Considerations**

The courts have recognized that students attending public schools have property rights and liberty interests in educational matters that trigger due process protections in situations involving high stakes testing as a measure of student performance and/or accountability. Due process
in high-stakes testing has been defined by the courts to include four specific rights:

- The right to expect that the test is an accurate and appropriately used assessment tool;
- The right to be tested only on the curriculum that has been taught;
- The right to expect the test and the students’ level of preparation for the test to meet professionally accepted standards; and
- The right to demand that the state prove that the students required to take the test have been adequately prepared to take the test and that they have a reasonable chance to pass the test (Debra P. v. Turlington, 1984).

High-stakes testing is the use of performance data from a standardized, usually criterion-referenced, test to make critical decisions about students, often regarding issues of promotion, retention, grouping, and/or graduation. There is a great deal of public and political support for the use of high-stakes tests as the sole measure of accountability, but the rights of students who take the test must still be acknowledged and protected (Elul, 1999).

The courts require the protection of due process for students involved in matters related to discipline, but do not extend the protection of due process to situations in which academic decisions are challenged. However, two significant Supreme Court decisions have recognized due process rights in the academic context.

In The Board of Curators of the University of Missouri v. Horowitz (1978) and The Regents of the University of Michigan v. Ewing (1985), the court acknowledged that the students (plaintiffs) had liberty and property interests sufficient to trigger due process protection. This position was clarified by Justice Stevens, when in writing for the court in Missouri v. Horowitz he stated "that federal courts can review an academic decision of a public educational institution under a substantive due process standard" (Missouri v. Horowitz, 1978). Such an acknowledgement by the court is significant since it sets a precedent upon
which other challenges might proceed. For students seeking substantive due process review of high-stakes academic accountability policies or procedures, Missouri v. Horowitz offers considerable hope.

Significant hope is also offered by the due process analysis of standardized testing in Debra P. v. Turlington (1984). In this Florida case, students challenged a state law that required them to earn a particular score on a standardized test in order to earn a high school diploma. While the court was very careful not to venture into the area of educational policy determination, the court noted its obligation to ensure that the state exercise its control over education "in a manner consistent with the mandates of the United States Constitution" (Debra P. v. Turlington, 1984.). The student rights described by the Court in Debra P. v. Turlington are exactly those summarized at the beginning of this section.

Proponents of high-stakes testing have consistently tried to limit the requirements in the Debra P. case to educational situations involving the denial of a high school diploma. However, the door is now open and other challenges about academic decisions made at both the elementary and secondary school levels may well find their way through it into the courts (Dyson, 2002).

Equal Protection and Disparate Impact. Courts considering educational practices under the Equal Protection Clause of the Fourteenth Amendment have consistently found no violation if only a discriminatory effect exists and no discriminatory intent is found (Elul, 1999). This judicial trend is clear in Larry P. v. Riles (1985). In this case, the plaintiff challenged the California schools’ practice of using IQ tests to place students in remedial programs. The use of such tests had a racially disparate impact since it resulted in the placement of substantially more children of color into these remedial programs than their white counterparts. The Ninth Circuit Court dismissed the Equal Protection claim because of the absence of discriminatory intent as required in a previous and significant judicial ruling (Washington v. Davis, 1986).

Challenging student accountability practices under the Equal Protection Clause has failed to provide petitioners with relief. This situation is
highlighted in *Eric v. (Causby) Johnston County* (1997) and *GI Forum, Image de Tejas v. Texas Educational Agency* (1997), two noteworthy federal court challenges to student accountability programs. These two cases clarify the reasoning behind challenges based on the Equal Protection Clause of the Fourteenth Amendment and race-based classification issues.

In *Eric v. (Causby) Johnston County*, the plaintiff challenged the school district’s accountability system, which required students in grades three through eight who do not receive a designated score on a standardized test to be retained regardless of their academic performance throughout the preceding school year. In addition, the accountability regulations denied course credit to students in grades nine through twelve who failed to achieve a score of at least 70 on a mandated final exam, whether or not they had achieved passing grades in that course throughout the year. Students who failed the test were required to retake it after participating in a mandated remedial program.

The complaint centered around the plaintiff’s assertion that during the first two years of this accountability program’s implementation, minority students failed at twice the rate of white students at all grade levels tested. The plaintiff alleged that this school-based accountability system violated student due process and the Equal Protection Clause under the Fourteenth Amendment of the United States Constitution. The plaintiff also argued that the policy underlying the accountability system violated Title VI of the Civil Rights Act of 1964. The court denied the plaintiff’s motion for a preliminary injunction without deciding the merits of the case. In denying the request for a preliminary injunction, the court held that “federal courts have no business substituting their judgment for that of the local school board when it comes to qualitative standards for promotion.” This matter was finally settled by the parties without further judicial review. The quiet settlement included an agreement to make substantial changes in the requirements of the school district’s accountability system which ultimately resulted in fewer failures (Dyson, 2002).
In *GI Forum, Image de Tejas v. Texas Educational Agency*, the plaintiffs challenge that the statewide accountability program adversely affects students of color since the diploma denial rate for students of color due to failure to pass an exit exam was twice that for white students. The plaintiffs alleged violations of due process and the Equal Protection Clause of the Fourteenth Amendment of the United States Constitution, and violations of Title VI of the Civil Rights Act of 1964. The court upheld the state’s position and allowed the use of an exit exam for graduation, denying the disparate impact challenge.

Some in the legal field feel that challenges like those in *Eric* and *GI Forum* have little chance of succeeding (Braceras, 2002). On the other hand, others believe that of all the issues raised in these cases, the equal protection challenge may hold the most promise (Elul, 1999). However, such a challenge would require the plaintiffs to successfully argue that school accountability policies violate the Equal Protection Clause in two ways: by burdening a fundamental right and by creating a race-based classification. The logic here is straightforward but not without distractions. If a state policy, act, or law burdens a fundamental right or creates a race-based classification, it would be subject to strict scrutiny and perhaps found unconstitutional. Unfortunately, this premise is weak since education has not yet been recognized by the courts as a fundamental right secured by the Constitution. Also, asserting that school accountability policies are unconstitutional because they create a race-based classification would be difficult. Even though race-based classifications are subject to strict scrutiny under the Equal Protection Clause, the Supreme Court’s findings in such cases have consistently required a showing of discriminatory intent. Proving such intent and demonstrating a resulting disparate impact remains very difficult. This difficulty is crystallized in the Supreme Court’s decision in *Alexander v. Sandoval* (2001), in which the Court rejected a disparate impact challenge to a state driver’s license exam that was offered only in English, holding that Title VI does not provide a private right of action for claims of disparate impact.
Others remain undaunted and feel that there is hope for such legal challenges. David Troutt (2001) offers his opinion in an article in the *New York Times*:

Nevertheless, the disparate impact regulations under Title VI remain in force, thus allowing the federal government to withdraw funding from institutions (like school districts) that utilize tests or policies that result in disproportionate racial or ethnic outcomes. Moreover, supporters of the disparate impact theory may still proceed under 42 U.S.C.1983. In addition, supporters of the disparate impact approach have called upon Congress to amend Title VI to permit explicitly private law suits against federally funded institutions that maintain policies or practices that affect some racial or ethnic group differently than others (including schools that use standardized graduation exams) (Troutt, *The New York Times*, 2001).

The Validity of Educational Measures and Practices. There is little doubt that teachers and administrators have limited knowledge of the varied and complex legal implications of school policies and practices related to student accountability. However, teachers are called upon to implement accountability measures that include the often unsupported use of high-stakes testing.

Clearly, many in the profession who are best able to comment on how tests should be developed and used consistently caution that it is difficult, perhaps impossible, to capture the depth and breadth of a student’s knowledge via the use of a single assessment instrument (Haney, 2000). In fact, the use of multiple measures to make a decision is statistically, logically, and strongly encouraged by the Joint Committee on Standards for Educational and Psychological Testing (1999). Specifically, Standard 13.6 states, “In educational settings, a decision or characterization that will have a major impact on a student should not be made on the basis of a single test score. Other relevant information should be taken into account, if it will enhance the overall validity of the decision” (Joint Commission, 1999). The issue is not whether students and schools should be held accountable, but, instead, the use of an invalid assessment (testing) practice (Gallagher, 2000).
In some school districts, rigorous external accountability requirements have preceded efforts to provide more equal educational resources to schools. Needless to say, the lack of such resources has diminished the capacity of the school system to help all students meet more challenging standards. This issue might provide fertile ground for viable legal challenges, if state revenues and fiscal support for public schools plummet.

As important as these issues are, they are summarily disregarded by policymakers who extol the virtues of accountability systems that fail to meet the standards posited by prominent and knowledgeable professionals in the field of assessment and testing. Many issues must be carefully considered in the development of any accountability system. These issues should be explored in the light of test theory and practice, the proper use of statistics, and the legal context in which such accountability systems must ultimately operate. Unfortunately, the individuals responsible for the institution of accountability programs based on high-stakes tests are not necessarily familiar with the insight and information available in publications such as "The Use of Tests as Part of High-Stakes Decision-Making for Students: A Resource Guide for Educators and Policy-Makers" (2000), produced by the Office of Civil Rights (OCR).

Surprisingly, the guidelines and standards noted in such informative reports and studies have not dramatically influenced the actual assessment practices in most educational accountability systems, which continue to base educational decision-making on a single assessment tool. Sadly, highly informed professional organizations within the testing community offer standards that specify “best practice,” but no mechanism of enforcement. The legal community offers a mechanism of enforcement, but is reluctant to tread on the authority of those responsible for determining educational policy. This reluctance is obvious in court decisions such as *Castaneda v. Pickard* (1981). In *Castaneda*, where teachers were sent into bilingual classrooms without knowing enough Spanish to teach their students, the court emphasized that it had reached its decision in the case without unduly imposing
its educational values and theories in educational decisions generally reserved to the state or local school authorities.

Psychometricians and statisticians firmly assert that test validity, reliability, and fairness are issues of “best practice” that can legitimately be determined by adhering to principles included in test theory and by using meaningful statistical analysis. Meeting the standards of best practice encouraged by these groups should rely not on the whim of any political body but on the court’s acknowledgement of and respect for commonly understood best and most appropriate assessment practices.

Conclusion

If local school-based accountability programs are to include tests as gatekeepers to significant educational school opportunities, then the schools in which these systems are to operate must be mindful of appropriate testing practices and the legal challenges that have been brought in response to high-stakes testing mandates. These legal challenges and the professional knowledge available about testing must guide educators in their efforts to serve all students better. To do any less is educationally inappropriate, ethically deficient, and legally unacceptable.

Implications for Further Study

It is true that the debate around high-stakes testing raises a number of legal, ethical, and educational issues. These issues are worthy of further investigation and discussion because they have a clear and important relationship to the development and future implementation of school-based accountability systems. School practitioners charged with the responsibility of developing internal (school-based) accountability systems would be wise to fully understand the implications of the issues raised in this project. Having a better understanding and knowledge of these matters will help school practitioners develop and implement a valid and useful accountability system within an appropriate ethical and legal framework.
References


Castañeda v. Pickard, 648 F. 2d 989, (5th Cir. 1981)


Data Sources for Teacher Development and School Improvement


Larry P. v. Riles, 793 F. 2d 969 (9th Circuit 1984)


Shapiro, J., & Stefkovich, J. (2001). Ethical leadership and decision making in education.


Data Sources for Teacher Development and School Improvement


Washington v. Davis, 528 U.S. 858.