

**Assessment, high stakes, and alternative visions:
Appropriate use of the right tools to leverage
improvement**

by

**Dan Laitsch
Assistant Professor
Simon Fraser University**

Education Policy Research Unit (EPRU)
College of Education
Division of Educational Leadership and Policy Studies
Box 872411
Arizona State University
Tempe, AZ 85287-2411

November 2006

EPRU | EDUCATION POLICY RESEARCH UNIT

EPSL-0611-222-EPRU
<http://edpolicylab.org>

This research was made possible by funding from the Great Lakes Center for Education Research and Practice.

Assessment, high stakes, and alternative visions: Appropriate use of the right tools to leverage improvement

Daniel A. Laitsch

Simon Fraser University

Background

The past twenty-five years of education reform efforts can perhaps best be characterized as the standards, assessment, and accountability movement. Since publication of the report *A Nation at Risk*, educators and policymakers nationwide have struggled to define academic performance standards and design related assessments. Most recently, in the belief that a reward and punishment system will stimulate systemic school improvement, policymakers have used student test performance as the basis for awarding or denying resources and recognition to schools, educators, and students—a system known as high-stakes assessment.

While such an accountability system may be sensible on its face, it does not account for multiple educational purposes or the complexity of assessment. Nor does it allow educators opportunity to use appropriate data thoughtfully to design comprehensive school reform. Instead, the stress on rewards and punishments based on test scores forces schools to consider the data generated as evaluative rather than as useful for informing instruction. The result is a system that appears coordinated, but results in a number of unintended—although not unpredictable—negative consequences.

This paper examines the historical roots of assessment and explores various practical and theoretical concerns related to high-stakes assessment. An overview of alternative reform and accountability models and a review of relevant research follows.

Accountability Systems: Stakeholders

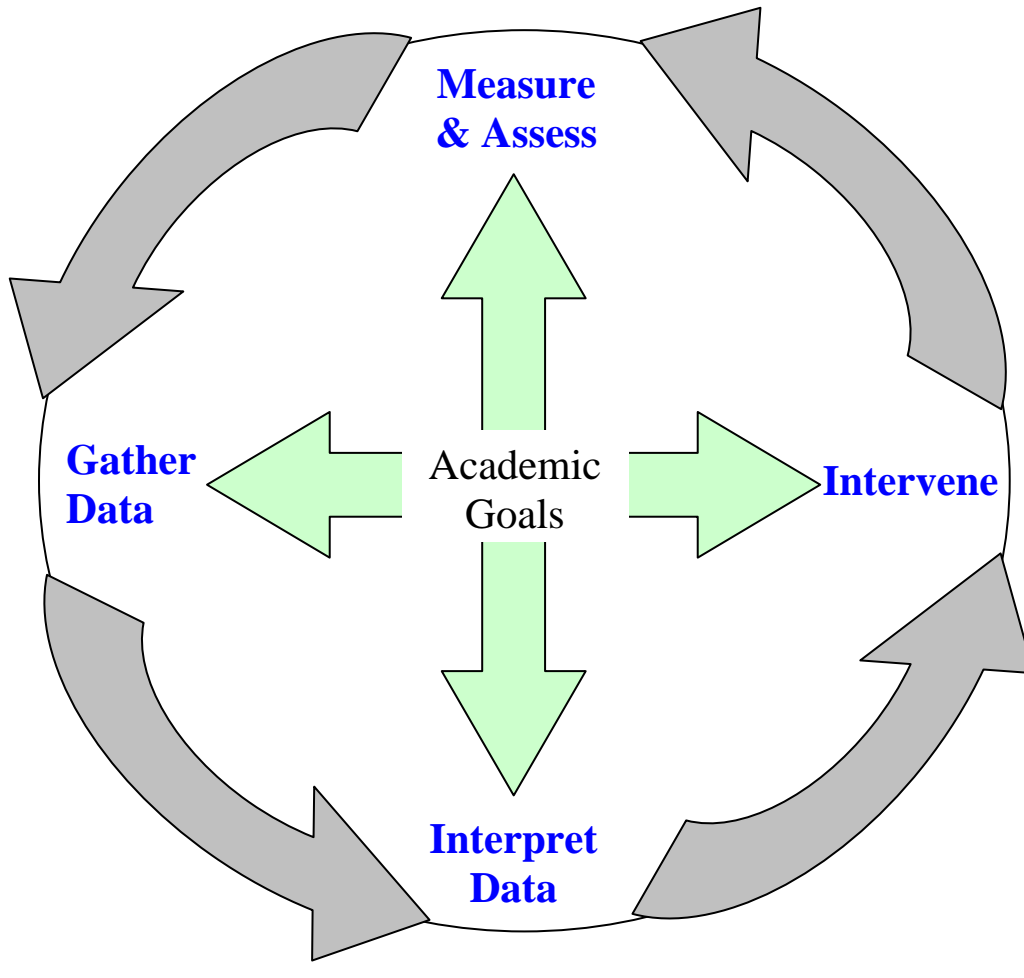
The current accountability system operates formally through governance structures at four levels: legislative, judicial, local, and school. In setting education policy, legislators generally exercise their authority through mandates that are primarily enforced through the allocation or withholding of resources. Because resource allocation is the primary tool for policymaker use in leveraging change, much of the current system rests in a framework focused on reward and sanction rather than capacity. As discussed later, this behaviorist policymaking model has serious implications for stakeholders. The judiciary is also active in education policymaking, and plays an important role in guiding and constraining legislative action based on civil rights and school finance litigation. However, the courts can only intervene when asked to so by an external audience. Local policy authority lies in school boards, which are directly responsible for managing the educational system and typically have hiring and other budgetary powers. The school level includes educators who generally enact mandates from the other levels; however, when it involves collective action, the local level can have substantial policy influence.

Two other interested groups can be loosely described as internal and external audiences. These groups have no formal authority, and rely instead on influencing others. The internal audience includes professional associations of educators, administrators, and policymakers; the external audience includes coalitions of parents, businesses, think tanks, and others interested but not directly engaged in education. Each

group has members with diverse ideas about public education's goals and about how to judge a school's effectiveness. In contrast, the current high-takes system assumes that it is self-evident that all schools should pursue increased test scores as their dominant goal and that those scores offer the most reliable evidence of how well a school is performing (see Figure 1).

However, the more complex reality is that various stakeholders push multiple goals for the public education system. These goals may include encouraging positive

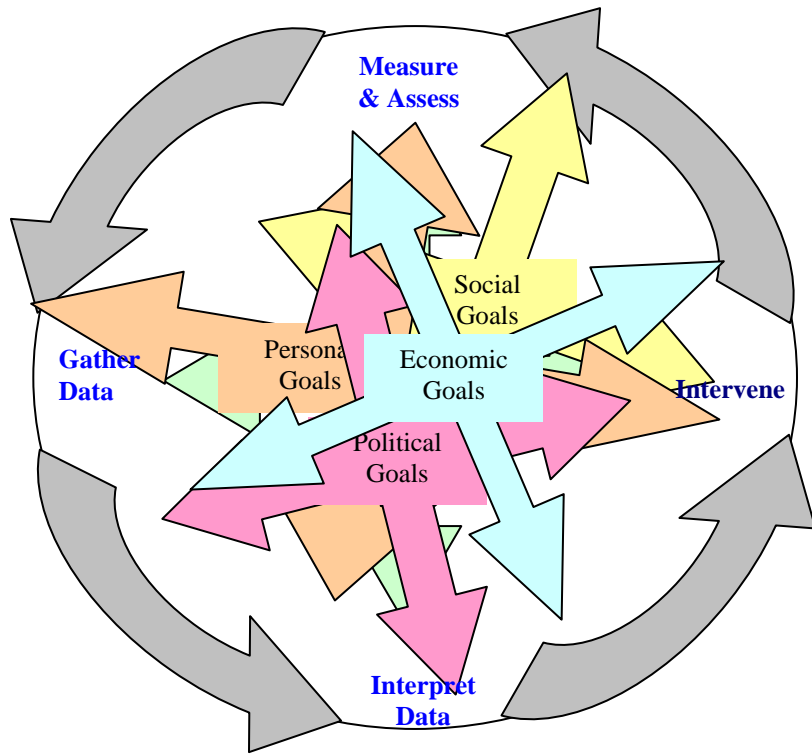
Figure 1: Accountability and change in education



socialization, strengthening common moral values, promoting civic engagement and equal opportunity, preparing responsible citizens, cultivating economic self-sufficiency, and promoting cultural unity.¹ When such multiple goals are considered, a much more complex and confused system of accountability emerges, highlighting competing interests and the difficulties of creating a unified system (see Figure 2).

In contrast to the presumed simplicity of the high-stakes model, the complex reality involves competing stakeholders seeking to exert their authority. Moreover, the goals

Figure 2: Competing goals for accountability and change



they seek are not only their own but also those of multiple audiences who have won their support. While academic proficiency is an important goal of schools, stakeholders continue to lobby for social, political and/or economic goals, creating tension throughout the system, affecting research design, and ultimately shaping education reform.

Educational Theory and Accountability

Behavioral Psychology and High-Stakes Testing

The high-stakes model is hierarchical, with each level of formal stakeholders seeking to influence others, usually below them. Thus, legislators try to shape schools through funding and policy, school administrators try to influence teachers' classroom activities, and teachers try to shape student behavior and performance. The high stakes imposed by the legislature begins this chain of action.

Absent from the process is any attention to the interest or motivation of those the legislature wishes to change, consistent with the behavioral psychology that supports the approach. Behavioral psychology asserts that people are conditioned to behave in specific ways through punishment and reinforcement. The expectation is that because students, teachers and administrators want such rewards as diplomas and salary increases, and to avoid such punishments as being retained or fired, they will work to increase academic achievement, provide better services to low-achieving students and populations, increase professional development in academic areas, and improve planning focused on the identified curriculum—irrespective of resources allocations or other external influences.^{2, 3}

However, such outcomes are less reliable than the theory suggests.⁴ For example, if policymakers use a high-stakes system to push a district to increase math test scores, they assume that the district will respond responsibly. It might, for example, adopt a new curriculum, provide teachers appropriate professional development and necessary class time to implement it, and develop student support systems like tutoring. However, a range of negative results is also possible. Teachers might decide to narrow their efforts

and teach only tested topics, or administrators may add instructional time to math by cutting back other, equally important classes. In both cases, scores rise. However, the second scenario demonstrates additional, unintended and negative consequences: students experience an impoverished academic experience. In effect, high-stakes systems may result in practitioners changing their behavior from what they consider ethical best practice to altered, undesirable behavior in order to achieve the mandated outcomes and avoid punitive consequences. Both theory and experience suggest that a high-stakes environment does in fact promote a wide variety of negative consequences, as will be detailed in a later segment of this brief.^{5, 6, 7, 8, 9, 10}

While policymakers can attempt to design high-stakes systems that minimize unintended consequences,¹¹ human behavior is not so easily managed. The entire high-stakes system depends on a series of questionable assumptions about behavior: that education professionals and systems will not pursue improvement unless coerced by punishments and rewards; that the system is in crisis and needs radical change that can only be driven externally; and, that improved academic achievement can be met without increasing system-wide capacity. If even one of these assumptions is unreliable,^{12, 13, 14, 15, 16} then the framework for the current accountability system begins to fail.¹⁷

One Alternative: Self-determination Theory

Behaviorism is not the only theory useful for developing an accountability system; in fact, others may provide a better base,¹⁸ in part by predicting and addressing the unintended consequences of high-stakes systems.¹⁹ Self-determination theory (SDT), one such alternative, emphasizes individual empowerment and control (rather than rewards and sanctions). Generally, SDT stresses the importance of offering appropriate

informational experiences aligned with achievable expectations—in this case, academic standards. The theory suggests that individuals respond to experience in one of three ways. They may embrace change when they feel an experience was informational; they may accept change with a weak commitment to it when they feel an experience was controlling; or, they may lose motivation to change and essentially give up if they feel the target goal of an experience was unattainable. The theory makes clear that no one is motivated by probable failure.²⁰ The current system under NCLB then may be unlikely to promote positive or lasting change, in part because the current high-stakes system promotes the expectation of failure. When NCLB was enacted, Congress set a goal that is arguably unattainable when it mandated a proficiency level of 100 percent, as almost any large system will contain enough variability to preclude 100 percent of its members performing at or above any specified level). In contrast, an accountability system using an SDT framework would shift the emphasis from external control (rewards and punishments) to internal control (empowered stakeholders) and capacity. SDT, in fact, aligns well with current organizational theory, which has de-emphasized competition and sanctions.^{21, 22}

Assessment: Practical Issues

While beyond the scope of this paper, the history of testing is rife with controversy and complexity. Despite creation of many widely used tests (such as the Stanford-Binet Intelligence Scale and the Scholastic Aptitude Test), researchers are still debating whether intelligence is a single entity²³ or whether there are in fact multiple types of intelligence.²⁴ Researchers are also trying to determine to what extent intelligence is genetic and to what extent it is influenced by environment.²⁵ Historically,

standardized intelligence testing has been used to justify racial and ethnic discrimination as well as to provide the foundation for the Eugenics movement.²⁶ In some schools, Eugenics policies were used to track students and allocate resources based on students' theoretical genetic worth and potential. Such abuse is not simply a historical issue—in 2004 almost 60,000 Tennessee voters cast ballots for James Hart, a candidate for U.S. Congress running on a Eugenics platform.^{27, 28, 29}

This brief overview is not meant to discredit current testing policies or the field of psychometrics, which has aggressively worked against such lamentable discrimination. It is, however, critical that stakeholders understand the history that makes many distrustful of standardized testing programs, as well as their potential for both intentional and unintentional misuse. To prevent abuse of testing programs, several professional testing organizations have jointly published *The Standards for Educational and Psychological Testing* to help guide responsible—and ethical—use of assessment.³⁰ A key recommendation is that test data not be used as the sole indicator in a high-stakes accountability system—precisely the current situation.

Assessment and Data

Of key importance are issues of validity (is the assessment measure the student knowledge or ability it is designed to measure), reliability (does the test generate consistent data about student knowledge or ability), and scoring (does the score accurately represent student ability—e.g. below proficient, proficient, or above proficient). Even well designed tests can only examine the performance of students on a small part of the entire curriculum and at only one point in time, meaning that most of what a student knows and is able to do remains unmeasured. As a result, high-stakes

accountability systems are generally not sophisticated enough to accommodate the different contexts across classrooms, schools and districts.

A simple example can be seen in the manner in which student turnover is handled. Many schools – particularly schools with a large enrollment of high poverty students – experience significant student mobility, with 30 percent or more of their students leaving or arriving in any given year (some schools have reported turnover in excess of 100%). In such cases, school effectiveness is confounded by prior experience—in other words, student outcomes are not the result of the current school and teachers, rather they are a reflection of prior educational experiences. The failure of current mechanisms to account for this type of variability in context is just one symptom of a larger problem in assuming standardized assessment outcomes to be a valid measure of student and school performance. Again, the problem is not the data, it is the way in which the data is used.

Another issue relates to the difficulty of determining who is responsible for the ethical use of assessments. Currently, the ethical responsibilities for appropriate design and application are spilt between test maker and test user.³¹ The test maker is responsible for designing a test that is reliable and a valid measure of the knowledge and skills that test purchasers have contracted for. The test purchasers are then responsible for the ethical application of the test, as well as the appropriate use of data gleaned from the test. This creates an unfortunate separation of ethical accountability, that allows test makers to develop, sell, and score assessments that are being misused, or used in questionable ways, by the purchasers.

Recent Developments in Accountability

The 2002 reauthorization of the Elementary and Secondary Education Act (ESEA), known as No Child Left Behind (NCLB), intended to improve multiple facets of education, but it focused primarily on academic achievement. Aggressive accountability language and strict enforcement of the regulations have created a national accountability system that uses federal funding as a high-stakes tool to leverage systemic change at the state, regional and local level. In fact, for the first time in the history of ESEA, the Department of Education has fined states that it deemed out of compliance with the law.³² Concurrently, resistance has grown and spawned several proposed adjustments or alternatives to the current system.

No Child Left Behind

Much has been written elsewhere about the intricacies of the NCLB accountability model,^{33, 34, 35} and only a brief summary is offered here.³⁶ As an assessment and accountability system, NCLB requires schools to attain 100 per cent student proficiency in math and literacy by the 2013-14 school year. En route, schools must demonstrate “adequate yearly progress” (AYP) by setting and attaining increasingly higher target goals. Improvement must also occur for every subgroup of students, including those from low socioeconomic backgrounds, major racial and ethnic groups, students with disabilities, and students with limited English proficiency. Schools that receive Title I funds and consistently fail to make adequate progress are then subject to a series of progressively harsher sanctions that range from allowing students to transfer to higher achieving schools and funding private tutoring to reconstitution, dismissal of staff, or even closure.

Alternative Models

Many states have considered forfeiting federal funds rather than adopting the NCLB accountability model.³⁷ However, most states and schools don't find it politically feasible to forfeit federal monies, even though they provide only a small percentage of state education expenditures.³⁸ Although most states have adopted the NCLB model, many stakeholders continue to seek alternatives. Proposals range from pragmatic models seeking modification of the NCLB system to theoretical models that involve more sweeping – and less specific – recommendations for change.

Legislative Models

Practical proposals focus on persuading authorities to strategically change the current system, assuming that it is unlikely that the existing system can be re-invented, and that instead its weakest aspects should be revised through legislative or regulatory change. However, these models generally continue to embrace high-stakes testing, and so retain many of that system's shortcomings.

Growth Model

The growth model is perhaps the most modest proposed change, and so it is one of the most viable proposals. In fact, the Department of Education has already agreed to pilot tests in two states.^{39, 40}

Growth models attempt to measure how much growth individual students exhibit from one year to the next; the score of an individual student in a subject one year is compared to his or her score in the subject a year later. This contrasts with the current system, which compares performance of different student groups; the scores of students in grade three this year, for example, are compared to the scores of students in grade three

last year. A growth model assumes that a more accurate reflection of improvement can be gained from comparing an individual student against him- or herself.

FairTest proposal

The National Center for Fair and Open Testing (FairTest) is a nonprofit group that advocates for more fair, transparent, and beneficial assessment.⁴¹ In 2004, FairTest helped orchestrate the development and publication of several recommended changes to NCLB that are backed by 70 national professional associations, including most major education associations.⁴² The proposal's 14 recommendations include ensuring that assessments align with state standards; meet professional and technical standards; are of adequate technical quality; provide measures of performance that assess higher order thinking skills; and provide diagnostic information that can be used to improve teaching and learning. Additionally, the recommendations urge using data only in valid and reliable ways, and easing the testing burden by limiting annual testing requirements.

Judicial models and adequacy

The judiciary also contributes to education reform through litigation focusing on school funding and equity. Recent “adequacy” lawsuits are not traditional reform efforts but they have potential to leverage great change. Such lawsuits challenge state education financing systems and try to establish that state elected leaders are accountable for providing adequate resources to schools. Historically, nearly every state has experienced legal challenges to its funding system, most often focused on equitable funding across a state.⁴³ Now, however, many school finance lawsuits focus on adequacy—that is, on whether state systems provide adequate resources for students to achieve the mandated

state standards.⁴⁴ For example, adequacy research indicates that the cost of bringing children from high-poverty backgrounds to proficiency on state assessments may be 35 to 100 per cent more than the cost for more advantaged students; English Language Learners may require additional resources of as much as 100 per cent.⁴⁵ Those promoting adequacy models want such differentials reflected in funding.

Adequacy lawsuits have been extremely successful. While states won approximately two-thirds of equity challenges, those challenging state finance systems on adequacy grounds are winning about two-thirds of their cases. Just recently, for example, courts ordered New York state to increase its annual funding for New York City schools by at least \$4.7 billion, as well as to invest an additional \$9.2 billion in school facilities.⁴⁶

Equal Protection

Earlier court challenges to high-stakes accountability, based largely on civil rights arguments, have evolved to include challenges based on equal protection grounds. These lawsuits claim that failure to provide adequate support has denied equal educational opportunity to poor and minority students and English Language Learners, all of whom have been disparately affected by high-stakes tests. This area is beyond the scope of this paper, but may prove significant. While the decision was ultimately overturned, a California judge initially suspended the state's high-stakes sanction that would deny diplomas to almost 50,000 students who had not passed the state's exit exam.⁴⁷

Audience Models

Because of the power of legislators and the judiciary to shape the current system, other external and internal audiences continue trying to persuade them to intervene on their behalf. Chief among the alternatives being promoted or resisted by various groups

are models that would establish teaching as a profession; address the health and welfare of students more generally; and introduce market mechanisms into the system.

Professional Model

Educators strongly support professional accountability models, asserting that education is a complex endeavor requiring a specialized body of knowledge and skills. These models incorporate self-regulation by the profession, based on the assumption that non-practitioners lack the necessary knowledge for assessing expertise.

Several professional organizations representing teachers, administrators, and researchers have been working to design such models, as have standards and accrediting bodies. Generally, professional models require that practitioners learn specialized knowledge and skills, engage in clinical practice, and after entering the profession, participate in professional development to remain current in their fields.⁴⁸ The models require that before working in a school, candidates must obtain certification or licensure, renewable every few years. Most models also include opportunities for advancement and professional development. In these models, members of the professional community are responsible for upholding ethical standards, disciplining members of the profession when necessary, and protecting the rights and welfare of their charges, the students.

Components of the professional model are currently in place in every state, although implementation is fragmented at best and strongly opposed by deregulation advocates. Every state licenses its teachers, and most require preparation requiring specific coursework as well as clinical experience; most also require that licenses be renewed, and many recognize advanced certification offered by the National Board for Professional Teaching Standards. To be sure that institutions preparing teachers offer

appropriate educational experiences, two nationally recognized organizations accredit teacher preparation programs, and a handful of states have enacted professional standards boards (PSB) to provide a mechanism for self-governance.

Despite the presence of many of these professionalism components, however, there has been little success in aligning them and protecting the integrity of each mechanism. Alternative certification programs and emergency licensure allow individuals to bypass the professional route, and state policymakers have been reluctant to relinquish their legislative authority and control of education to PSBs.

Coordinated School Health Model

Coordinated school health programs (CSHP) offer schools and communities the opportunity to shift focus of change efforts to the broader health of educational systems. By focusing on a wide variety of indicators, CSHP approaches broaden the dialogue and take the accountability focus off of testing academics and reframe it to support for broader assessment of multiple school and community variables, and targeted capacity building and improvement, rather than accountability and sanction.

CSHP models come from public health, but a coalition of internal audiences now supports them. They are based on the idea that neither students nor teachers can perform well if students are unhealthy, or forced to exist in an unhealthy environment.⁴⁹ CSHP models define health broadly, to include the mental, physical, and intellectual health of students, educators, and families, and supporting healthy systems and communities. A CSHP has the potential to make the current test-based system more reliable by making it less likely test data will be significantly skewed by disparities in student health.

Currently there are a variety of tools available to assess the health of schools that honor the voice of all education stakeholders. These include survey and assessment tools that yield prioritized lists of for creating a more healthy community. The Centers for Disease Control and Prevention is a strong advocate for CSHPs and has developed one such tool,⁵⁰ as has the Association for Supervision and Curriculum Development.⁵¹ Canadian and Australian governments have also supported development of such tools.⁵² Combined with district-level guidance and standards, the CSHP approach can help build a system focused on the needs of the whole child, not simply on his or her test scores.

Market Models

The models described above all propose accountability based on the assumption that the current system of public education will remain dominant. In contrast, the market model would fundamentally alter school governance and funding structures by introducing market forces that, theoretically, will improve school performance.

The market model includes two forms of choice: the regulated market, which allows for choice among public and private schools, and public school choice, which allows for choice among only public schools. Both models theorize that as consumers choose the best schools for their children, poor schools will go out of business and good schools will flourish and that competition for students will stimulate improvement. The government's role in these systems would be sharply curtailed, limited to basic monitoring and oversight.⁵³ Most commonly in market models, parents or guardians receive tuition vouchers, paid for by public funds, which they use to pay all or part of the tuition at the school of their choice, within the limits of the particular model.

Market model voucher programs are controversial for many reasons. Among the most common are concerns that competition won't adequately ensure quality and that private purposes for education, such as religious indoctrination, will supersede public purposes. There are also constitutional concerns about state funding of religious schools, and state constitutional responsibilities for maintaining a public school system. Still, various voucher programs are in place in multiple cities and states, and Congress recently established a national voucher program for students displaced by Hurricane Katrina.⁵⁴

In a public school choice system, market forces are commonly introduced through open enrollment, magnet schools, or charter schools, with charter schools being particularly evident. Government specified entities (school boards, for example) approve proposals for charter schools, which are then funded through public education funds. Charters are an attempted compromise that allows for semi-public control of the schools through the chartering agency, yet introduces market forces by allowing parents and guardians to choose schools for their children.

Educational Management Organizations (EMOs) are another market model. These are professional management companies contracted to manage one or more schools; a major component of their current business is running charter schools. Perhaps the best known example is Edison Schools, which now manages schools and educational programs in 25 states as well as in some other countries.⁵⁵

Those who promote a move toward market systems generally assume that parents, guardians and students will choose a school based on its success in helping students attain academic goals. As Figure 2 (page 5) indicates, however, many stakeholders may be

more concerned with other educational goals, so that relying on choice may not promote the academic improvements that market advocates seek.

Available Data

Available data on the effectiveness of the various accountability mechanisms is mixed. There is much research on high-stakes assessment, but it needs to be considered cautiously. Data for the other models is scarce because their implementation is limited. The notable exception is data on market accountability models, but this area also merits caution when considering research reports.

High-Stakes Testing and Student Learning

No Child Left Behind

The Department of Education has recently claimed that No Child Left Behind (NCLB) reforms have resulted in landmark improvements in student achievement, in both reading and math, and for African Americans as well as Hispanics, as evidenced by the National Assessment of Educational Progress (NAEP).⁵⁶ Despite these broad claims of success, other researchers have argued that there was no change in average reading scores on the NAEP between 2002 (before NCLB effects would likely be measured) and 2006, and only a small increase in math scores.⁵⁷ Thus, the Administration's claims of great success may be overstated.

Other Analysis

Other studies offer a wider variety of findings. Some have found that high-stakes tests are important in increasing student achievement,^{58, 59} while others have found significant problems with high-stakes models.^{60, 61, 62, 63} Still others have found while

that high-stakes assessments result in no increase in learning (and sometimes indicate a decrease), they do produce increased student retention and dropout rates.

A study based on data from the 1992-96 and 1996-2000 NAEP cohorts, estimated academic growth during the time states were actively creating high-stakes assessment systems.⁶⁴ Researchers compared scores in 28 states having “consequential accountability” with scores in 14 states having only “report card” accountability. They found that students in states with consequential systems showed more academic gain than those in report card states—but also that high-stakes systems widened the gap in scores between Black and White students.⁶⁵ Another study that examined the correlation between high and low stakes tests in two states and a number of localities found that gains on the high-stakes tests were accompanied by gains on low stakes tests. However, as will be discussed later, the correlation may be skewed by the results from one anomalous state.⁶⁶

Other studies have associated high-stakes exams with more negative outcomes. A study which explored the impact of high school exit exams, found that students in states with such mandated exams had lower graduation rates and poorer SAT scores than their peers in non-exit-exam states. In addition, students in minority groups and from high poverty backgrounds were disproportionately affected. Interestingly, special education students showed increased graduation rates, possibly because some were excluded from exit exam requirements as part of their IEP.⁶⁷

Research by Walter Haney on the effects of the Texas Assessment of Academic Skills (TAAS), has stimulated many similar studies that have generally or partially confirmed his initial results.^{68, 69, 70} Haney’s findings relative to high-stakes assessment

indicated that schools devoted extensive time and resources specifically to test preparation. While such test rehearsal produced large gains on the TAAS, similar gains were not seen on other tests, suggesting little or no increase in actual learning. Haney hypothesized that as much as half the TAAS gains may have been caused by retention of students in pre-test grades and increased dropout rates, which both decrease the number of low scoring students in the test population. Similarly, the number of special education students nearly doubled in four years, again lowering the number of low achieving test-takers. Because of such negative effects on students, Haney concluded that the Texas experience demonstrated that testing needed to be properly reconceptualized “as a source of potentially useful information to inform human judgment, and not as a cudgel for implementing education policy.”⁷¹

Others studies have also failed to establish a correlation between high-stakes accountability and improvement on other assessments, which would be expected if students were experiencing real learning gains. One study compared student performance on state assessments in 18 high-stakes states with their performance on other assessments (SAT, ACT, NAEP, and the AP exams). The results indicated little relationship between performance on state tests and the other measures, and the researched concluded that the high-stakes scores do not “appear valid as indicators of genuine learning.”

One study, interestingly, demonstrated that a high-stakes-low stress format avoided many of the negative effects seen in other states.⁷² At the time of the research, the state used two tests: one designed to measure complex knowledge and skills, and one designed to assess minimum competency, which students had to pass in order to graduate. Precisely because it was minimum competency, however, it had a pass rate of

over 99 percent, making it a low-stress assessment that did not appear to corrupt instruction striving for higher goals.

Unintended Consequences

Any system is likely to have unintended outcomes, but a high-stakes system may be particularly likely to exacerbate undesirable and unintended results.

Researchers have identified an observable score pattern across tests in high-stakes systems. When a new test first appears, scores drop precipitously, then rise significantly, and finally plateau—until another new test appears, when scores again drop and the pattern repeats itself.⁷³ When new tests appear, teachers and students lack familiarity with their content and structure, and instruction may not align well with the assessment. Scores fall. Then, significant gains follow as everyone becomes familiar with the new test and teachers realign instruction. Eventually however, scores reach a ceiling of near-maximum potential and further gains become more difficult. This pattern suggests that the extended drive for an ongoing linear increase in scores may be unrealistic, since ceilings do occur.⁷⁴

An emphasis on test scores predictably triggers the cycle, which occurs as teachers focus instruction on the tested areas. Strategies for improving scores may be pedagogically sound (aligning instruction with standards) or pedagogically unsound (coaching to narrow skills sets or knowledge domains). Systematic ways to affect scores can also be found, as when schools find ways to remove low scoring students from the pool of test takers. Again, these exclusions may be pedagogically sound (excluding special education students whose IEPs demonstrate that such assessment would be inappropriate) or pedagogically unsound (transferring students to other schools,

misidentifying students as eligible for special education, or retaining low scoring students prior to testing grades).

Narrowed Curriculum and Instructional Strategies

Policymakers see high-stakes assessments as a tool to pressure schools and educators to focus on particular areas they think important. However, it is exceptionally difficult to force change, and especially so with the fairly simplistic tool of high-stakes tests. Recent research suggests that in high-stakes environments, teachers and schools alter instruction and curricula to emphasize the band of disciplines tested—currently literacy and numeracy.⁷⁵ A Center on Education Policy survey found that more than 70 percent of respondents reported reducing attention to other subjects in order to spend more time on reading and math in elementary school while struggling students were given as much as doubled time on tested subjects, and in some cases missed other classes completely.⁷⁶ A similar effect was documented in another North Carolina.⁷⁷ Thus, the price for higher test scores may be significant restriction of a student’s overall educational experience.

Research also suggests that teachers alter instruction and use teaching techniques (such as lecturing on discrete facts) that, while effective for raising standardized test scores, do not necessarily help students develop higher-order thinking and problem solving skills.^{78, 79} On one survey 75 percent of responding teachers indicated that they felt high-stakes tests were forcing them to teach in pedagogically unsound ways.⁸⁰

Efforts to Bypass the Tests

Other responses to the need to raise test scores push political and ethical boundaries. In some states and Canadian provinces, educators have urged parents and

students to boycott required tests,^{81, 82, 83} while in England, teachers narrowly avoided a boycott of national curriculum tests on a voting technicality.⁸⁴ In some schools and districts, teachers and administrators desperate to raise test scores have gone to the extreme of changing scores or providing students with answers.^{85, 86, 87} In other cases students have been “pushed out,” or counseled to leave the school, in an effort to remove traditionally low-scoring students from the pool of test takers.^{88, 89, 90} In addition, there has been a considerable, and suspect, increase in special education referrals and identification during the high-stakes assessment era.^{91, 92}

Differential Impacts on Student Subgroups

Minority, English Language Learner, high poverty, and special education students are particularly vulnerable to the negative outcomes from high-stakes assessments. NCLB mandates that these student subgroups be tested and their scores reported, with significant consequences for schools. When schools fail to make adequate yearly progress due to the performance of various subgroups, potential penalties for the school create a tension between inclusion and achievement that may lead to a series of consequences, particularly for special education students.⁹³ Positive outcomes include increased opportunities for special education students to have experiences and instruction more comparable to that of their peers; however, these positive effects depend on appropriate accommodations (such as test scores not precluding graduation). When states offer little flexibility regarding test-based retention, significant negative outcomes have resulted, including student retention based on a single test score, a greater presence of overage students in lower grades, insufficient remediation, and increased pressure on students to pass the test.⁹⁴

Similar benefits and drawbacks have been observed for students who are English Language Learners (ELLs) and who have had less access than special education students to appropriate accommodations. As some researchers have noted, schools risk being labeled as failing under NCLB simply due to high enrollments of ELL students, whose test performance cannot fairly be expected to equal that of native language speakers. States, however, have been slow to adjust accountability systems to appropriately and fairly assess ELL students.⁹⁵

Studies have also found disparate impacts on Black and Hispanic students. Although there is significant variation across states, strong high-stakes accountability systems appear to correlate negatively with promotion rates and positively with retention rates (a strong dropout indicator).^{96, 97}

Reallocation of Services to Students Closest to the “Cut” Scores

Another consequence of high-stakes-assessment systems has been a reallocation of resources away from the highest and lowest achieving students to those students closest to the cut scores on high-stakes examinations.⁹⁸ Students likely to pass the tests easily are left to manage on their own, as are students who are so far from passing the test that it is exceptionally unlikely that they will succeed. Instead, the limited resources available are strategically redirected to the students just under the mandated accountability thresholds, where they are likely to have the greatest impact on school effectiveness ratings. While effective at improving school ratings in the short term, such “educational triage” carries with it serious ethical implications.

Negative Impact of Testing Errors

Researchers have found that human error in testing programs occurs during all phases of testing – from design and administration to scoring and reporting – and that such errors can have a serious consequences in a high-stakes environment. In one year alone researchers found that human error caused 50 separate errors throughout the assessment process, affecting hundreds of thousands of students nationally: 50 students were wrongly denied graduation; 8,668 students were needlessly required to attend summer school; and 257,000 students were misclassified as limited-English-proficient.⁹⁹ In January of 2003, an ETS scoring error resulted in more than 4,000 teacher candidates being incorrectly failed on their certification tests, resulting in an \$11.1 million settlement of various lawsuits.^{100, 101} Shortly after this settlement, ETS announced another scoring error—this time on the SAT, affecting more than 5,000 students. For some students, scores were reported as much as 450 points lower than their actual scores on a 2,500-point scale. Because colleges had made admissions decisions by the time the error was announced, it is likely that many of these students were unjustly eliminated from consideration for admission and scholarships at competitive institutions.¹⁰² In Alabama, a Harcourt test was incorrectly scored for 25,000 students, affecting 600 schools under NCLB guidelines.¹⁰³ Connecticut fined Harcourt when it reported incorrect reading scores for 355 students across 51 districts.¹⁰⁴

Such scoring errors in a high-stakes system can have a lifelong impact on students who may be inappropriately denied diplomas, forced into unnecessary course work, or inappropriately retained in a lower grade. Such consequences partly explain why research groups have opposed the use of a single test in high-stakes systems.

Other Models

There is relatively little data available regarding other non-test-based accountability mechanisms, either because they are not yet used extensively or because they involve only proposed modifications not yet adopted. However, some discussions and reviews of alternative proposals illuminate their complexities and possible outcomes.

Growth Model

Researchers, value-added practitioners, and testing companies have offered critical reviews of value-added growth models, resulting in serious concerns about “vertical scaling,”¹⁰⁵ a fundamental component of most growth models. Such scaling involves comparing test scores vertically—or across grades. The practice rests on several assumptions, including that the content of different tests is sufficiently similar to allow for comparison. This is clearly inappropriate for subjects like science, where students may study biology one year and physics the next. Even in the less stratified subjects of mathematics and reading serious concerns remain.¹⁰⁶ While vertical scaling may reflect improvement in common content over two grade levels, important grade-specific content would not be measured. For example, a test based primarily on geometry will not capture what a student has learned in a previous grade about algebra. Additionally, sophisticated statistical analyses by multiple researchers suggests that such scaling is useful only in adjacent grades, and even then may result in both under- and over-estimates of school-wide performance.¹⁰⁷ Such issues suggest that great caution be taken when considering the use of value added models. Other reviews, which have focused specifically on the use of value added methods for evaluating teachers rather than systems can generally be

summarized as concluding that value added assessments should never be used as a sole indicator of quality.^{108, 109, 110, 111}

A major problem inherent in any value added systems is that success is defined by the outcome measured, in this case good test scores—suggesting that the experience is irrelevant as long as gains appear. Reviews of the most extensive such system in place, the Tennessee Value Added Assessment System (TVAAS), identify other concerns. For example, since the system is norm referenced it compares teachers to other teachers in the field rather than to an externally identified standard for “good teaching.”¹¹² In other words, the system ensures there will always be “good” and “bad” teachers, even if all students were meeting state standards. Moreover, researchers have argued that it does not account for teacher effects outside of test scores, and that it fails to account for student demographics,^{113, 114, 115} although TVAAS affiliated researchers claim the system does adequately account for the influence of student race, socioeconomic status (SES), and prior achievement.¹¹⁶ Effects of TVAAS have not been externally validated, largely because data and technical documentation have not been released, as is customary in research efforts.

An additional issue here is that although even test companies like ETS and Harcourt urge caution and argue against misuse of value-added measurements, it is policy makers who ultimately control implementation. The proposed changes to NCLB would likely replace current assessment requirements with value-added assessment, leaving a single measure as the primary indicator for achieving AYP and therefore continuing the inappropriate practice of relying on a single indicator to determine major consequences for students and schools.

Adequacy Lawsuits

Research into the adequacy lawsuits is limited. Findings from work on New Jersey's adequacy-inspired effort to reform its funding system confirm implementation can be difficult, particularly when it requires program restructuring and resource reallocation. Obstacles include stakeholders clinging to entrenched practices and professionals lacking appropriate skills and professional development opportunities.¹¹⁷

Despite such complications, it is well documented that a primary problem is adequate support for schools.¹¹⁸ Research confirms both that resources matter and that they are unequally distributed, with gaps among districts as large as \$11,000 per student.¹¹⁹ Another study found that 22 states provide fewer dollars per student in high-poverty districts; 28 states provide fewer resources to students in high-minority districts; and, when the increased costs of educating high-poverty students is taken into account, 29 states have an average funding gap of over \$1,000 per student.¹²⁰

FairTest model

There is no research on the proposed FairTest model, although it includes components of the growth and professional models discussed in other sections. However, for this model, viability is a major concern. It has been endorsed by some 70 educational and civil rights organizations, including such high profile education organizations as the National Education Association (NEA), the American Federation of Teachers (AFT), and the National School Boards Association (NSBA). While such support sounds impressive, the political reality is that the education establishment has lost significant influence in policy arenas as Republicans have developed their own education agenda and advocacy groups.^{121, 122} Moreover, the Democratic education agenda – long supported by teacher

unions and other education advocacy groups – has splintered in the face of increasing support for testing and market ideologies.¹²³ New policy groups, such as the Progressive Policy Institute and the New Democrats, have frequently aligned themselves with conservative market-oriented think-tanks like the Fordham Foundation and the American Enterprise Institute. The education profession’s current isolation, reduced access, and eroded influence suggest that there may be little hope for the FairTest proposal.

Professional Model

While a significant body of research suggests that the professional model of teaching can have significant positive effects on students,^{124, 125, 126} proponents of market models both for schools and for teacher preparation have generated research supporting an agenda of deprofessionalization.^{127, 128, 129} Under the Bush administration, government reports on teacher quality have strongly promoted alternate teacher certification methods and removing state regulation of teacher preparation and licensure. Over the objections of peer reviewers, the Administration used discretionary grants to heavily invest in two new organizations dedicated to promoting the deregulation agenda.¹³⁰

In this political climate, the professional model of teaching – from teacher certification to professional standards boards – is vulnerable. Even where professional standards are supported many basic principles are compromised.¹³¹ While the research suggests that professional models could help improve schools, there are significant questions about the political will to enact this type of reform.

Coordinated School Health Model

While an extensive literature argues for CSHPs, there has been relatively little evaluation research on their implementation and outcomes. A Florida study found that CSHPs could be implemented and sustained for a relatively small investment (\$15,000 per school over 8 schools, plus personnel); however, the researchers did not find correlative improvements in student academic performance.¹³² A possibility here, however, is that the time span of the study was too short for gains to become evident. In Ohio, a survey of 225 districts revealed that while many districts were familiar with the model, few had coordinated programs in place that addressed all eight CSHP components.¹³³

Market Model

Researchers and activists have long debated the effects of various school governance systems on student achievement. Much of the market reform rhetoric assumes that privately run schools can effect higher student achievement at a similar or lower cost than is possible in public schools. However, researchers have argued that once students' background traits are taken into account, the difference in achievement between public and private school students disappears. In other words, variation in student performance has little to do with how a school is funded and governed; rather, it is a product of other variables, including student socioeconomic status (SES), parental education, community support, peer group characteristics—and school structures, including school and class size. Two carefully conducted studies released in the summer of 2006 found little or no difference between the achievement of students in public and private systems once such traits were statistically controlled.^{134, 135}

Researchers looking specifically at voucher programs have found mixed outcomes. Depending on the design and author, studies have found positive effects,¹³⁶
¹³⁷ ¹³⁸ ¹³⁹ negative or null effects,^{140, 141, 142, 143} or mixed effects.^{144, 145} While studies conducted by market advocates have consistently found significant achievement gains, studies by independent researchers have failed to produce similar results. This lack of consensus has resulted in significant politicization of the debate surrounding the impact of vouchers on student achievement and accusations of bias and poor scholarship. Advocacy research conducted by organizations with a defined pro- or anti-voucher agenda has further clouded the picture. The end result is that many studies of voucher programs, irrespective of methodological quality, attract criticism from researchers and writers who hold views in opposition to those found in the reports.

The case of New York City's privately funded voucher program offers a useful illustration of the complexity of voucher research. One study found significant positive gains by African Americans who used vouchers; however, shortly after release of the findings, one of the researchers involved took the unusual step of retracting the claims of success, noting that the findings were skewed by an unusually strong performance by one subgroup within the study.¹⁴⁶ Moreover, analysis by different researchers in 2003 found that a large amount of data had not been used in the original analysis, and that when it was included, the voucher effects virtually disappeared.¹⁴⁷ Similar debates have occurred regarding the effectiveness of voucher programs in Florida,^{148, 149, 150} Ohio, and Wisconsin.^{151, 152} At best, the research into the effectiveness of vouchers as a reform mechanism is inconsistent, showing different results (which are generally small) for different grades, subjects, and localities.

The research on the effectiveness of charter schools is similarly unclear (some studies show positive outcomes,¹⁵³ some negative,¹⁵⁴ some mixed¹⁵⁵), although a number of recent and high profile studies have shown charter schools to be less effective than comparable public schools.^{156, 157} As with vouchers, much of the research is supported and published by advocacy groups with a particular perspective on the issue.

The effectiveness of EMOs generally mirror the effectiveness of the particular model school involved; that is EMO-managed charter schools perform similarly to charter schools generally. In the most comprehensive evaluation of an EMO, researchers found that students in Edison schools generally do as well as or better than those in conventional public schools, but gains for specific schools vary significantly depending upon implementation and duration of the program.¹⁵⁸

It should be noted that much of the research assessing the merits of privatization does not take into account the fact that whatever form schools take, they tend to approach education similarly. Schools are generally organized the same, have access to the similar funding, teach similar curricula, are evaluated using the same tests, draw from the same pool of teachers who are trained in a similar manner to meet the same standards, and so on. Because of these similarities, variable outcomes may be more closely tied to areas of greater difference, such as student, family, and community characteristics. It is possible that this situation contributes to the lack of clear findings in much of this research.

Evaluation of Available Data—Sources

Another characteristic of the current situation that surely confuses the research picture is that those with particular political agendas want data and use it to promote their goals, especially when those goals may not enjoy academic consensus or support. This

prompts agenda-driven research with different researchers drawing significantly different conclusions from the same data. The need for research supporting a politically desirable conclusion often overshadows concern for relying on only high quality, peer-reviewed studies. The result is an educational debate occurring in a maelstrom of conflicting studies and partisan research.

Therefore, research consumers should examine research critically to determine the possible political agenda of the researcher, the quality of the research, and the veracity of the recommendations. It is beyond the scope of this paper to provide such analysis for all of the studies referenced above, but comments are in order concerning research from two key sources—the Department of Education and market-oriented think tanks.

U.S. Department of Education

It is important to note that the Department of Education is a cabinet level office within the President's administration, and its leadership works to further the President's political agenda. Hence, it is obviously a political entity. Despite Congressional mandates that education research be, "objective, secular, neutral, and non-ideological and free of partisan political influence and racial, cultural, gender, or regional bias," President Bush has reserved the right to construe such provisions as advisory,¹⁵⁹ to establish a research agenda, and to suppress the publication of selected findings not helpful to political goals.¹⁶⁰ A brief review of the Administration's approach to educational research offers useful insight into the reliability of claims that NCLB has been strikingly successful and uncovers a pattern of past abuse of the public trust.

Propaganda, Peer Review, and Issue Advocacy

The U.S. General Accountability Office (GAO) has investigated accusations of ethical breaches at the Department of Education, and identified a number of illegal and unethical practices. The GAO determined that the Administration violated prohibitions against “covert propaganda” when it contracted with Armstrong Williams, an African-American spokesperson and columnist, to promote the No Child Left Behind Act to audiences without disclosing that he was being paid by the Department of Education.¹⁶¹ The GAO also ruled that the Department engaged in covert propaganda when it distributed fake TV and print news stories without disclosing its involvement.^{162, 163}

The GAO has further determined that the Department ignored the advice of peer reviewers when it funded an unsolicited proposal to establish the National Council on Teacher Quality, a group that advocates for the deregulation of teaching. The GAO found four other grants that the Department awarded in 2001 that had not been recommended for funding by any of the peer reviewers.¹⁶⁴ Republican and Democratic leaders in the Senate have recently asked the GAO to investigate the \$1 billion Reading First Program to determine whether senior administration officials illegally influenced what reading books schools could buy, whether specific publishers got preferential treatment, and whether states were pressured to buy the materials in order to get federal grant dollars.¹⁶⁵

In addition to such explicitly illegal activities, mandated reports to Congress on Title II of the Higher Education Act have been heavily criticized as lacking scholarship and as being propaganda for administration priorities.^{166, 167, 168, 169} In fact, the Department of its own admission ignored much of the peer-reviewed research on teacher

quality, stating that it used studies “which officials found to be more credible than many journals requiring peer review.”¹⁷⁰

Moreover, the Department of Education and the National Center for Education Statistics (NCES) delayed for almost a year the release of data from the National Assessment of Educational Progress (NAEP) that showed students at charter schools – an administration priority – performed poorly on the NAEP when compared with their public school counterparts.¹⁷¹ The Department also refused to release a research study it commissioned that concluded bilingual education programs are superior to all-English immersion programs—which are a conservative policy priority.¹⁷² Such a consistent pattern of abuse should make readers wary of Department of Education publications, particularly when the research findings support Administration political priorities.

Market-oriented Think Tanks

As noted above, political activists and advocacy organizations actively co-opt research related to market reforms in education. Of particular note is the sophisticated network of fiscally conservative think tanks and policy centers that have created a parallel system of ideologically-oriented scholarship that bypasses traditional scientific structures and quality control mechanisms.^{173, 174, 175} These entities operate under the assumption that “the war of ideas has largely been captured by free market proponents and that the remaining challenge centers on the implementation of sound policy.”¹⁷⁶ Their goal is not the development of public policy based on current research and information—rather these organizations seek to advance market ideology irrespective of available evidence.¹⁷⁷ Readers of research published by these groups (including among others the Heritage Foundation, the Fordham Foundation, the Center for Education

Reform, and the Hoover Institution) should be aware of these fiscal and ideological relationships as they evaluate the veracity of the data, methods, and interpretations of the authors—particularly if their outcomes consistently support market oriented reforms and policies.

Recommendations

If there is one overarching finding of the work discussed in this paper, it is that there is no silver bullet for school improvement. Education systems are complex and serve multiple purposes that are approached with a diversity of programs and services. Simplistic accountability mechanisms focused on single outcome measures and faulty assumptions about the behavior of individuals and systems cannot adequately assess the work of teachers, schools, and students, or provide sufficient information for policy decisions. Despite the general dearth of conclusive research regarding what works in educational accountability, the research does suggest some general directions that hold promise for substantive and lasting improvement.

It is recommended that policy makers:

1. Refocus reform emphasis to include building school capacity as well imposing professional accountability.
2. Abandon high-stakes accountability mechanisms, which produce not only questionable improvement in student learning but also unintended, significant negative consequences.
3. Align new assessment systems with professional guidelines for ethical use of assessment data.

4. Broaden the methods of data collection to better evaluate the multiple purposes of education.

Notes & References

- ¹ Cuban, L. (2006, February). Why can't schools be like businesses? *The School Administrator* 63(2). Retrieved May 20, 2006 from <http://www.aasa.org/publications/saarticledetail.cfm?ItemNumber=5212&snItemNumber=950&tnItemNumber=951>.
- ² Defur, S. (2002). Education Reform, High stakes Assessment, and Students with Disabilities. *Remedial & Special Education*, 23(4), 203. Retrieved Wednesday, April 12, 2006 from the Academic Search Elite database.
- ³ Stecher, B., & Barron, S. (2001). Unintended consequences of test-based accountability when testing in "milepost" grades. *Educational Assessment* 7(4). pp. 259-281. Retrieved Wednesday, April 12, 2006 from the Academic Search Elite database.
- ⁴ Ryan, R.M. & Brown, K.W. (2005). Legislating competence: The motivational impact of high stakes testing as an educational reform. In C. Dweck & A. E. Elliot (Eds.) *Handbook of Competence*. New York: Guilford Press.
- ⁵ Abedi, J. (2004). The No Child Left Behind Act and English Language Learners: Assessment and accountability issues. *Educational Researcher* 33(1). pp. 4-14.
- ⁶ Rhodes, K., & Madaus, G. (2003). *Errors in standardized tests: A systemic problem*. (National Board on Educational Testing and Public Policy Monograph). Boston, MA: Boston College, Lynch School of Education. Retrieved June 3, 2005, from <http://www.bc.edu/research/nbetpp/statements/M1N4.pdf>.
- ⁷ Booher-Jennings, J. (2005, Summer). Below the bubble: "Educational Triage" and the Texas Accountability System. *American Educational Research Journal* 42 (2). pp. 231-268.
- ⁸ Defur, S. (2002). Education Reform, High stakes Assessment, and Students with Disabilities. *Remedial & Special Education*, 23(4), 203. Retrieved Wednesday, April 12, 2006 from the Academic Search Elite database.
- ⁹ Stecher, B.M. & Hamilton, L.S. (2002, Spring). Putting theory to the test: Systems of "educational accountability" should be held accountable. *Rand Review* 26(1). pp. 17-23. Retrieved April 9, 2006 from <http://www.rand.org/publications/randreview/issues/rr.04.02/rr.spring2002.pdf>
- ¹⁰ Stecher, B., & Barron, S. (2001). Unintended Consequences of Test-Based Accountability When Testing in "Milepost" Grades. *Educational Assessment*, 7(4), 259-281. Retrieved Wednesday, April 12, 2006 from the Academic Search Elite database.
- ¹¹ Yeh, S. S. (2005). Limiting the unintended consequences of high stakes testing. *Education Policy Analysis Archives* 13(43). Retrieved April 3, 2006 from <http://epaa.asu.edu/epaa/v13n43/>.
- ¹² Berliner, D. & Biddle, B. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. Reading, MA: Addison-Wesley.
- ¹³ Levin, B. (1998, August 20). Criticizing the schools: Then and now. *Education Policy Analysis Archives* 6(16). Retrieved May 20, 2006 from <http://epaa.asu.edu/epaa/v6n16.html>.
- ¹⁴ Cizek, Gregory J. (June, 1999). Give Us This day our daily dread: Manufacturing crises in education. *Phi Delta Kappan*. Retrieved May 20, 2006 from <http://www.pdkintl.org/kappan/kciz9906.htm>.
- ¹⁵ Rothstein, R. (1998) *The way we were? Debunking the myth of America's declining schools*. Washington, DC: The Brookings Institute.
- ¹⁶ Cuban, L. (2006, February). Why can't schools be like businesses? *The School Administrator* 63(2). Retrieved May 20, 2006 from

- <http://www.aasa.org/publications/saarticledetail.cfm?ItemNumber=5212&snItemNumber=950&tnItemNumber=951>.
- ¹⁷ There is evidence that at least two of these assumptions are wrong: the system may not be in crisis, and the motivation for change is not simply academic improvement. The assumption that educators within the system don't want to change is also questionable, and most of the professional groups in the internal audience have invested considerable resources in the professional development and in efforts to engage legislators and the judiciary in change efforts.
- ¹⁸ Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55, 68-78. Retrieved November 23, 2004 from <http://www.psych.rochester.edu/SDT/documents/2000RyanDeciSDT.pdf>
- ¹⁹ Stecher, B.M. & Hamilton, L.S. (2002, Spring). Putting theory to the test: Systems of "educational accountability" should be held accountable. *Rand Review* 26(1). pp. 17-23. Retrieved April 9, 2006 from <http://www.rand.org/publications/randreview/issues/rr.04.02/rr.spring2002.pdf>
- ²⁰ Self-Determination Theory (2006). The High stakes Testing Controversy: "Higher Standards" Can Prompt Poorer Education. *Controversies*. Retrieved April 3, 2006 from http://www.psych.rochester.edu/SDT/cont_testing.html.
- ²¹ Wang, X. H. & Yang, B. Z. (2003). Why Competition may Discourage Students from Learning? A Behavioral Economic Analysis. *Education Economics* 11(2).
- ²² Crow, R. (1995). Institutionalized competition and its effects on teamwork. *Journal for Quality & Participation* 18(3). pp. 46-53.
- ²³ Gottfredson, L. (1998, January 1). *The general intelligence factor*. Scientific American Presents: Exploring Intelligence. See: <http://www.sciamedigital.com/index.cfm>. Retrieved April 7, 2006 from <http://www.psych.utoronto.ca/~reingold/courses/intelligence/cache/1198gottfred.html>
- ²⁴ Checkley, K. (1997, September). The First Seven...and the Eighth: A Conversation with Howard Gardner. *Educational Leadership*. Alexandria, VA: Association for Supervision and Curriculum Development.
- ²⁵ Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science* 14(6). Retrieved September 20, 2004 from <http://www.blackwell-synergy.com/links/doi/10.1046/j.0956-7976.2003.psci.1475.x.abs>.
- ²⁶ Stoskopf, A. (1999). Echoes of a forgotten past: Eugenics, testing, and education reform. *Rethinking Schools Online* 13(3). Retrieved April 8, 2006 from http://www.rethinkingschools.org/archive/13_03/eugenic.shtml.
- ²⁷ Associated Press. (2004). *Elections 2004: Tennessee*. Washington Post. Retrieved April 8, 2006 from <http://www.washingtonpost.com/wp-srv/elections/2004/tn/>
- ²⁸ Hart, J. (2004). James Hart for Congress. Retrieved April 7, 2006 from <http://www.jameshartforcongress.com/>.
- ²⁹ Washington Post (2004). *Politics: Elections 2004*. Washington, DC: Author. Retrieved April 12 from <http://www.washingtonpost.com/wp-srv/elections/2004/candidates/22307/>.
- ³⁰ American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: AERA.
- ³¹ Megargee, E. I. (2000). *Testing*. Encyclopedia of psychology, Vol. 8. Kazdin, A. E. (Ed). Washington, DC: American Psychological Association. pp. 47-52.
- ³² Clark, K. (2005). SN CJ Spotlight: States rebel against No Child Left Behind. *StateNet Capitol Journal* 13(21). Retrieved April 4, 2006 from http://www.legislate.com/capitol_journal/06-13-2005.

- ³³ Linn, R.L., Baker, E.L., & Betebenner, D.W. (2006). Accountability Systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher* 31(6). pp. 3–16. Retrieved March 21, 2006 from <http://www.learningexperts.com/McQuillan/NCLBEdResearcher093002.pdf>.
- ³⁴ Erpenbach, W.J., Forte-Fast, E., & Potts, A. (2003). Statewide Educational Accountability Under NCLB: The State Collaborative on Assessment and Student Standards (SCASS), Accountability Systems And Reporting (ASR) Consortium. Washington, DC: the Council of Chief State School Officers. Retrieved April 3, 2006 from <http://www.ccsso.org/content/pdfs/StatewideEducationalAccountabilityUnderNCLB.pdf>.
- ³⁵ Laitsch, D. (2003, October). No Child Left Behind: Reflections on implementation: Two years and counting. ASCD Infobrief. Washington, DC: Association for Supervision and Curriculum Development. Retrieved April 8, 2006 from <http://www.ascd.org> Web site.
- ³⁶ Laitsch, D. (2006). Search of the ECS NCLB Database (National Grid). Denver, CO: Education Commission of the States. Retrieved April 11, 2006 from <http://nclb2.ecs.org/NCLBSURVEY/nclb.aspx?Target=NGDetails>.
- ³⁷ Peterson, K. (2005, July 07). No letup in unrest over Bush school law. *Stateline.org*. Retrieved October 31, 2006 from <http://www.stateline.org/live/ViewPage.action?siteNodeId=136&languageId=1&contentId=41610>
- ³⁸ National Priorities Project. (2006). Underfunding for Title I. Northampton, MA: Author. Retrieved March 26, 2006 from http://www.nationalpriorities.org/index.php?option=com_content&task=view&id=83&Itemid=61.
- ³⁹ U.S. Department of Education. (2006). *Elementary & Secondary Education: Growth models*. Washington, DC: Author. Retrieved April 2, 2006 from <http://www.ed.gov/admins/lead/account/growthmodel/index.html>.
- ⁴⁰ U.S. Department of Education. (2006, May 17). Press release: Secretary Spellings approves Tennessee and North Carolina growth model pilots for 2005-2006; 6 other states to get early consideration for 2006-2007. Washington, DC: Author. Retrieved May 20, 2006 from <http://www.ed.gov/news/pressreleases/2006/05/05172006a.html>.
- ⁴¹ National Center for Fair and Open Testing (undated). About Fairtest. Cambridge, MA: Author. Retrieved April 10, 2006 from <http://www.fairtest.org/Who%20We%20Are.html>.
- ⁴² National Center for Fair and Open Testing (2004, October 21). *Joint organizational statement on No Child Left Behind (NCLB) Act*. Cambridge, MA: Author. Retrieved April 10, 2006 from <http://www.fairtest.org/joint%20statement%20civil%20rights%20grps%2010-21-04.html>.
- ⁴³ Hunter, M. (2005). *Litigations challenging constitutionality of K-12 funding in the 50 states*. New York: Campaign for Fiscal Equity. Retrieved April 12, from <http://www.schoolfunding.info/litigation/In-Process%20Litigations.pdf>.
- ⁴⁴ Hunter, M. (2005). *Litigation*. New York: Campaign for Fiscal Equity. Retrieved April 12, 2006 from <http://www.schoolfunding.info/litigation/litigation.php3>.
- ⁴⁵ Baker, B. (2005, Winter). The emerging shape of educational adequacy: From theoretical assumptions to empirical evidence. *Journal of Education Finance* 30(3). pp. 259-287.
- ⁴⁶ Hunter, M. (2006, April 10). *Litigation: New York*. New York: Campaign for Fiscal Equity. Retrieved October 31, 2006 from http://www.schoolfunding.info/states/ny/lit_ny.php3
- ⁴⁷ Asimov, N. (2006, May 9). Judge says California exit exam is unfair. San Francisco Chronicle. Retrieved May 20, 2006 from <http://www.sfgate.com/cgi-bin/article.cgi%3Ffile%3D/c/a/2006/05/09/MNGSVIO7NII.DTL&cid=0>.

- ⁴⁸ Education Commission of the States. (2003). Next-generation models of education accountability: The teacher professionalism model. Denver, CO: Author. Retrieved March 30, 2006 from <http://www.ecs.org/clearinghouse/46/30/4630.doc>.
- ⁴⁹ Hanson, T. L., & Austin G. A. (2003). Are student health risks and low resilience assets an impediment to the academic progress of schools? *California Healthy Kids Survey: Factsheet 3*. Los Alamitos, CA: WestEd
- ⁵⁰ Centers for Disease Control. (2005). Healthy youth! Welcome to the school health index. Silver Spring, MD: Author. Retrieved April 2, 2006 from <http://apps.nccd.cdc.gov/shi/default.aspx>.
- ⁵¹ Association for Supervision and Curriculum Development (2006). Creating a healthy school using the healthy school report card: An ASCD action tool. Alexandria, VA: Author. Information retrieved April 5, 2006 from <http://www.ascd.org/portal/site/ascd/menuitem.ab8d1b51d1c866f8f7378b10d3108a0c/>.
- ⁵² Vamos, S. (2006). Comprehensive School Health: Monitoring, Assessing, and Evaluating Process and Outcomes. Workshop presented to the annual meeting of the Alberta Coalition for Healthy School Communities.
- ⁵³ Adams, J.E. & Hill, P.T. (2003). Next-generation models of education accountability: The regulated market model. Denver, CO: Education Commission of the States. Retrieved March 30, 2006 from <http://www.ecs.org/clearinghouse/46/32/4632.doc>.
- ⁵⁴ Devany, L. (2006, June 15). Congress Approves Additional \$235 Million for Katrina Vouchers. *Press release*. Alliance for School Choice. Retrieved October 31, 2006 from <http://www.allianceforschoolchoice.org/more.aspx?IITypeID=3&IIID=2701>
- ⁵⁵ Molnar, A., Garcia, D.R., Bertlett, M., & O'Neill, A. (2006, May). Profiles of For-Profit Education Management Organizations: Eighth Annual Report, 2005-2006. Tempe, AZ: Arizona State University Education Policy Studies Laboratory, Commercialism in Education Research Unit. Retrieved June 12, 2006 from <http://www.asu.edu/educ/eps/CERU/Documents/EPsL-0605-104-CERU.pdf>.
- ⁵⁶ U.S. Department of Education. (2006, April). Overview: No Child Left Behind Act is working. Washington, DC: Author. Retrieved April 17, 2006 from <http://www.ed.gov/nclb/overview/importance/nclbworking.html>.
- ⁵⁷ Center on Education Policy. (2006). *From the Capital to the classroom: Year 4 of the No Child Left Behind Act*. Washington, DC: Author. Retrieved April 12, 2006 from <http://www.ctredpol.org/nclb/Year4/CEP-NCLB-Report-4.pdf>.
- ⁵⁸ Rosenshine, B. (2003, August 4). High stakes testing: Another analysis. *Education Policy Analysis Archives*, 11(24). Retrieved May 18, 2006 from <http://epaa.asu.edu/epaa/v11n24/>.
- ⁵⁹ Greene, J., Winters, M., & Forster, G. (2004). Testing high stakes tests: Can we believe the results of accountability tests? *Teachers College Record* 106(6). pp. 1124-1144. Retrieved April 18, 2006 from <http://www.tcrecord.org/content.asp?contentid=11568>.
- ⁶⁰ Marchant, G. J. & Paulson, S. E. (2005, January 21). The relationship of high school graduation exams to graduation rates and SAT scores. *Education Policy Analysis Archives* 13(6). Retrieved April 15, 2006 from <http://epaa.asu.edu/epaa/v13n6/>.
- ⁶¹ Amrein, A.L. & Berliner, D.C. (2002, March 28). High stakes testing, uncertainty, and student learning *Education Policy Analysis Archives*, 10(18). Retrieved April 14, 2006 from <http://epaa.asu.edu/epaa/v10n18/>.
- ⁶² Amrein, A.L. & Berliner, D.C. (2003, August 4). Re-analysis of NAEP math and reading Scores in states with and without high stakes tests: Response to Rosenshine. *Education Policy Analysis Archives*, 11(25). Retrieved April 14, 2006 from <http://epaa.asu.edu/epaa/v11n25>.

- ⁶³ Haney, W. (2000, August 19). The myth of the Texas miracle in education. *Education Policy Analysis Archives* 8(41). Retrieved April 4, 2006 from <http://epaa.asu.edu/epaa/v8n41/>.
- ⁶⁴ Hanushek, E.A. & Raymond, M.E. (2004, September). Does school accountability lead to improved student performance? *NBER Working Papers*. Cambridge, MA: National Bureau of Economic Research. Retrieved April 18, 2006 from <http://edpro.stanford.edu/Hanushek/admin/pages/files/uploads/accountability.jpam.journal.pdf&e=9797>.
- ⁶⁵ Hanushek, E.A. & Raymond, M.E. (2004, September). Does school accountability lead to improved student performance? *NBER Working Papers*. Cambridge, MA: National Bureau of Economic Research. pp. 22. Retrieved April 18, 2006 from <http://edpro.stanford.edu/Hanushek/admin/pages/files/uploads/accountability.jpam.journal.pdf&e=9797>.
- ⁶⁶ Greene, J., Winters, M., & Forster, G. (2004). Testing high stakes tests: Can we believe the results of accountability tests? *Teachers College Record* 106(6). pp. 1124-1144. Retrieved April 18, 2006 from <http://www.tcrecord.org/content.asp?contentid=11568>.
- ⁶⁷ Marchant, G. J. & Paulson, S. E. (2005, January 21). The relationship of high school graduation exams to graduation rates and SAT scores. *Education Policy Analysis Archives* 13(6). Retrieved April 15, 2006 from <http://epaa.asu.edu/epaa/v13n6/>.
- ⁶⁸ Kellow, J.T. & Willson, V.L. (2001). Consequences of (mis)use of the Texas assessment of academic skills (TAAS) for high stakes decisions: A comment on Haney and the Texas miracle in education. *Practical Assessment, Research & Evaluation*, 7(24). Retrieved April 9, 2006 from <http://PAREonline.net/getvn.asp?v=7&n=24>.
- ⁶⁹ Klein, S.P., Hamilton, L.S, McCaffrey, D.F., Stecher, B.M. (2000). What do test scores in Texas tell us? *Education Policy Analysis Archives*, 8(48). Retrieved April 18, 2006 from <http://epaa.asu.edu/epaa/v8n49/>.
- ⁷⁰ Camilli, G. (2000). Texas Gains on NAEP: Points of Light? *Education Policy Analysis Archives*, 8(42). Retrieved April 18, 2006 from <http://epaa.asu.edu/epaa/v8n42.html>.
- ⁷¹ Haney, W. (2000, August 19). The myth of the Texas miracle in education. *Education Policy Analysis Archives* 8(41). Retrieved April 4, 2006 from <http://epaa.asu.edu/epaa/v8n41/>
- ⁷² Yeh, S. S. (2005). Limiting the unintended consequences of high stakes testing. *Education Policy Analysis Archives*, 13(43). Retrieved April 11, 2006 from <http://epaa.asu.edu/epaa/v13n43/>.
- ⁷³ Linn, R. (2000). Assessments and accountability. *Educational Researcher* 29 (2). pp. 4-16.
- ⁷⁴ Linn, R. (2000). Assessments and accountability. *Educational Researcher* 29 (2). pp. 4-16.
- ⁷⁵ Coutinho, D., Nartowicz, N. & Penabad, D. (2006). *Curriculum shifts in Vermont public schools: A survey of school superintendents*. Retrieved March 22, 2006 from <http://policyresearch.dartmouth.edu/assets/pdf/nclb.pdf>.
- ⁷⁶ Center on Education Policy. (2006). *From the Capital to the classroom: Year 4 of the No Child Left Behind Act*. Washington, DC: Author. Retrieved April 12, 2006 from <http://www.ctredpol.org/nclb/Year4/CEP-NCLB-Report-4.pdf>.
- ⁷⁷ Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*. 81. pg. 200.
- ⁷⁸ Berry, B., Turchi, L., Johnson, D., Hare, D., Owens, D., & Clements, S. (2003). *The impact of high stakes accountability on teachers' professional development: Evidence from the South*. Chapel Hill, NC: Southeast Center for Teaching Quality, Inc. Retrieved June 7, 2004, from http://www.teachingquality.org/resources/pdfs/Spencer_FinalReport.pdf.

- ⁷⁹ Popham, J. (2005, April/May). F for assessment. *Edutopia*. Retrieved March 23, 2006 from http://www.edutopia.org/magazine/edl/article.php?id=art_1267&issue=apr_05.
- ⁸⁰ Pedulla, (2003, November). State-Mandated Testing—What Do Teachers Think? *Educational Leadership* 61(3). pp. 42-46.
- ⁸¹ British Columbia Teachers' Federation. (2006). FSA testing can be harmful to students! Vancouver, BC: Author. Retrieved April 22, 2006 from <http://www.bctf.ca/fsa/brochure/view.pdf>.
- ⁸² St. John, K. (2001, May 4). Marin students boycott state test: Achievement test boycotted in Marin district—Schools left ineligible for state reward funds. *San Francisco Chronicle*. Retrieved April 20, 2006 from <http://www.sfgate.com/cgi-bin/article.cgi?f=/chronicle/archive/2001/05/04/MN231705.DTL>.
- ⁸³ Associated Press. (2001). Scarsdale students skip class to protest tests. *CNN Education*. Retrieved April 15, 2006 from <http://archives.cnn.com/2001/fyi/teachers.ednews/05/03/test.boycott.ap/>.
- ⁸⁴ BBC. (2003). Teachers reject boycott of tests: Teachers will not boycott next year's tests for seven and 11 year olds, avoiding a threatened confrontation with the government. *BBC News*. Retrieved April 3, 2006 from <http://news.bbc.co.uk/1/hi/education/3323803.stm>.
- ⁸⁵ Rodriguez, N. (1999, July 22). Cheating teacher skews schools' test scores. *Eagle-Tribune* (Lawrence, MA). Retrieved November 23, 2004 from http://www.eagletribune.com/news/stories/19990722/FP_002.htm.
- ⁸⁶ Jacob, B.A. & Leavitt, S.D. (2004). To catch a cheat. *Education Next* 4(1). Retrieved March 30, 2006 from <http://www.educationnext.org/20041/68.html>.
- ⁸⁷ Snell, L. (2005, June). How schools cheat: From underreporting violence to inflating graduation rates to fudging test scores, educators are lying to the American public. *Reason Online*. Reason Foundation. Retrieved April 2, 2006 from <http://www.reason.com/0506/fe.ls.how.shtml>.
- ⁸⁸ Lewin, T. & Medina, J. (2003, July 31). To cut failure rate, schools shed students. *New York Times*. Section A , pg. 1. Retrieval as of April 20, 2006 from <http://select.nytimes.com/gst/abstract.html?res=FA0B11FB3F5A0C728FDDAE0894DB404482>.
- ⁸⁹ Kelleher, M. (1999, June). Dropout rate climbs as schools dump truants. *Catalyst Chicago*. Retrieved March 31, 2006 from <http://www.catalyst-chicago.org/news/index.php?item=330&cat=23>.
- ⁹⁰ Darling-Hammond. (2004, Fall). Gross inequities: Confused priorities. *Connections* (11)1. Retrieved April 3 from https://www.publiceducation.org/pdf/Publications/Connections/Connections_Fall04.pdf.
- ⁹¹ Amrein, A.L. & Berliner, D.C. (2002, March 28). High stakes testing, uncertainty, and student learning *Education Policy Analysis Archives*, 10(18). Retrieved April 14, 2006 from <http://epaa.asu.edu/epaa/v10n18/>.
- ⁹² Haney, W. (2000, August 19). The myth of the Texas miracle in education. *Education Policy Analysis Archives* 8(41). Retrieved April 4, 2006 from <http://epaa.asu.edu/epaa/v8n41/>.
- ⁹³ Ysseldke, J., Nelson, R., Christenson, S., Johnson, D.R., Dennison, A., Triezenberg, H., Sharpe, M., & Hawes, M. (2006). What We Know and Need to Know about the Consequences of High stakes Testing. *Students with Disabilities Journal*, 71(1). pp. 75-95.
- ⁹⁴ Ysseldke, J., Nelson, R., Christenson, S., Johnson, D.R., Dennison, A., Triezenberg, H., Sharpe, M., & Hawes, M. (2006). What We Know and Need to Know about the Consequences of High stakes Testing. *Students with Disabilities Journal*, 71(1). pg. 89.
- ⁹⁵ Batt, L., Kim, J. & Sunderman, G. (2005, February). Limited English Proficient Students: Increased Accountability Under NCLB. Policy Brief. The Civil Rights Project At Harvard University. Retrieved March 23, 2006 from http://www.civilrightsproject.harvard.edu/research/esea/LEP_Policy_Brief.pdf.

- ⁹⁶ Carnoy, M. (2005). Have state accountability and high stakes tests influenced student progression rates in high school? *Educational Measurement: Issues and Practice* 24(4).
- ⁹⁷ Amrein, A.L. & Berliner, D. (2002, December). *An analysis of some unintended and negative consequences of high stakes testing*. Tempe, AZ: Education Policy Research Unit, Education Policy Studies Laboratory at Arizona State University. Retrieved April 20, 2006 from <http://www.asu.edu/educ/eps/EPRU/documents/EPST-0211-125-EPRU.pdf>.
- ⁹⁸ Booher-Jennings, J. (2005, Summer). Below the bubble: "Educational Triage" and the Texas Accountability System. *American Educational Research Journal*, 42(2). pp. 231-268.
- ⁹⁹ Rhodes, K., & Madaus, G. (2003). Errors in standardized tests: A systemic problem. (*National Board on Educational Testing and Public Policy Monograph*). Boston, MA: Boston College, Lynch School of Education. Retrieved March 3, 2006, from <http://www.bc.edu/research/nbetpp/statements/M1N4.pdf>.
- ¹⁰⁰ Clark, M. (2004, July 15). Some did pass teaching test: Essays were graded too harshly. *Cincinnati Enquirer*. Retrieved May 26, 2005, from http://www.enquirer.com/editions/2004/07/15/loc_teachertest15.html.
- ¹⁰¹ Arenson, K.W. (2006, March 15). Case involving errors in teacher test is settled. *New York Times*. Retrieved April 20, 2006 from <http://www.nytimes.com/2006/03/15/education/15sat.html>.
- ¹⁰² Newsinferno.com (2006, April 12). SAT scoring debacle prompts apology by College Board President. Retrieved April 20, 2006 from <http://www.newsinferno.com/archives/1074>.
- ¹⁰³ News 4. (2006, March 21). Alabama Testing Error. Dothan, AL: Author. Retrieved May 20, 2006 from <http://www.wtvnews4.com/schools/headlines/2502791.html>.
- ¹⁰⁴ Strenberg, B.J. (2006, February 17). *Reporting error on CAPT affects 51 school districts*. Hartford, CT: Connecticut Department of Education. Retrieved May 20, 2006 from <http://www.state.ct.us/sde/PressRoom/CAPTpressrelease.pdf>.
- ¹⁰⁵ Martineau, J.A. (2005, June 19). *Un-Distorting Measures of Growth: Alternatives to Traditional Vertical Scales*. Paper presented to the Council of Chief State School Officers 35th Annual National Conference on Large-Scale Assessment.
- ¹⁰⁶ Lissitz, R. W. & Huynh, H. (2003). Vertical equating for state assessments: issues and solutions in determination of adequate yearly progress and school accountability. *Practical Assessment, Research & Evaluation* 8(10). Retrieved June 25, 2006 from <http://PAREonline.net/getvn.asp?v=8&n=10>.
- ¹⁰⁷ Martineau, J.A. (2005, June 19). *Un-Distorting Measures of Growth: Alternatives to Traditional Vertical Scales*. Paper presented to the Council of Chief State School Officers 35th Annual National Conference on Large-Scale Assessment.
- ¹⁰⁸ McCaffrey, D.F., Lockwood, J.R., Koretz, D.M., & Hamilton, L.S. Evaluating Value-Added Models for Teacher Accountability. Santa Monica, CA: RAND. Retrieved April 20 from http://www.rand.org/pubs/monographs/2004/RAND_MG158.pdf&e=9797.
- ¹⁰⁹ Braun, H. I. (2005). Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models. *Perspective*. Princeton, NJ: Policy Information Center at ETS. Retrieved March 24, 2006 from <http://www.ets.org/Media/Research/pdf/PICVAM.pdf>.
- ¹¹⁰ Braun, H. I. (2005). Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models. *Perspective*. Princeton, NJ: Policy Information Center at ETS. Retrieved March 24, 2006 from <http://www.ets.org/Media/Research/pdf/PICVAM.pdf>.
- ¹¹¹ Harcourt Assessment, Inc. (2004, November). Value-added Assessment Systems. *Policy Report*. San Antonio, TX: Author. Retrieved April 20 from <http://harcourtassessment.com/hai/Images/pdf/assessmentReports/ValueAdded.pdf&e=9797>.

- ¹¹² Note that such a norm-referenced system ensures there will always be winners and losers—half the teachers will perform above the arbitrary norm, and half below it—even if all their students were meeting state standards.
- ¹¹³ Bracey, G. (2004, May). Serious questions about the Tennessee value-added assessment system. *Phi Delta Kappan* 85(9). pp. 716-717.
- ¹¹⁴ Bracey, G.W. (2006, February). Value-Added models, front and center. *Phi Delta Kappan* 87(6). p. 478-9.
- ¹¹⁵ Kupermintz, H. (2003). Value-added assessment of teachers. In A. Molnar (Ed.), *School reform proposals: The research evidence*. Tempe, AZ: Education Policy Studies Lab at Arizona State University. Retrieved March 28, 2006 from <http://www.asu.edu/educ/eps/EPRU/documents/EPRU%202002-101/Chapter%2011-Kupermintz-Final.htm>.
- ¹¹⁶ Sanders, W. L. (2004, October). Compounding Errors. Letter to the Editor. *Phi Delta Kappan* 86(2), p 174-175.
- ¹¹⁷ Erlichson, Goertz & Turnbull, 1999 and Erlichson & Goertz, 2001 as cited in Odden, A., Picus, L. O., & Fermanich, M. (2003, September 1). *An Evidence-Based Approach to School Finance Adequacy in Arkansas*. Final Report Prepared for the Arkansas Joint Committee on Educational Adequacy. North Hollywood, CA: Lawrence Picus and Associates.
- ¹¹⁸ Greenwald, R., Hedges, L., & Laine, R. (1996). The effect of school resources on student achievement. *Review of Educational Research*, 66(3), pp. 361–397.
- ¹¹⁹ Berliner, D., & Biddle, B. (2003). What research says about unequal funding for schools in America. *Policy Perspectives*. San Francisco, CA: WestEd. Retrieved April 12, 2006 from http://www.wested.org/online_pubs/pp-03-01.pdf.
- ¹²⁰ Carey, K. (2003, Fall). *The funding gap: Low-income and minority students still receive fewer dollars in many states*. Washington, DC: The Education Trust. Retrieved April 20, 2006 from <http://www2.edtrust.org/NR/rdonlyres/EE004C0A-D7B8-40A6-8A03-1F26B8228502/0/funding2003.pdf>.
- ¹²¹ DeBray, E. H. (2005, March 15). Partisanship and Ideology in the ESEA Reauthorization in the 106th and 107th Congresses. Paper presented for the Wilson Center for Scholars Symposium Panel on ESEA at 40. Retrieved April 22 from <http://www.wilsoncenter.org/events/docs/debray.doc&e=9797>.
- ¹²² Laitsch D., Heilman, E., & Shaker, P. (2002, December). Teacher Education, Pro-Market Policy and Advocacy Research. *Teaching Education* 13(3). pp. 251-271.
- ¹²³ DeBray, E. H. (2006). *Politics, ideology, and education: Federal policy during the Clinton and Bush administrations*. New York: Teachers College Press.
- ¹²⁴ Berry, B. (2001, May) No Shortcuts to Preparing Good Teachers. *Educational Leadership* 58(8). pp. 32-36.
- ¹²⁵ Darling-Hammond, L. (1999) *Teacher Quality and Student Achievement: A Review of State Policy Evidence*. Education Policy Analysis Archives. Volume 8 Number 1. Retrieved May 20, 2006 from <http://epaa.asu.edu/epaa/v8n1/>
- ¹²⁶ Darling-Hammond, L. & Youngs, P. (2002) *Defining “Highly qualified teachers”: What does “Scientifically based research” actually tell us?* Educational Researcher. Retrieved May 20, 2006 from http://www.aera.net/uploadedFiles/Journals_and_Publications/Journals/Educational_Researcher/3109/3109_ResNewsComment.pdf.
- ¹²⁷ Goldhaber, D., & Brewer, D. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129-145.

- ¹²⁸ Johnson, K. A. (2000). The Effects of Advanced Teacher Training in Education on Student Achievement. Washington, DC: Heritage Foundation. *Center for Data Analysis Report #00 09*. Retrieved May 20, 2006 from <http://www.heritage.org/Research/Education/CDA00-09.cfm>.
- ¹²⁹ Raymond, M., Fletcher, S. & Luque, J. (2001). *Teach For America: An evaluation of teacher differences and student outcomes in Houston, Texas*. Stanford, CA: Hoover Institute, Center for Research on Educational Outcomes (CREDO). Retrieved May 20, 2006 from <http://credo.stanford.edu/downloads/tfa.pdf>.
- ¹³⁰ Shaul, M.S. (2006, February). Discretionary grants: Further Tightening of Education's Procedures for Making Awards Could Improve Transparency and Accountability. *Report to the Ranking Minority Member, Education and the Workforce Committee, House of Representatives*. Washington, DC: United States Government Accountability Office. Report GAO-06-268. Retrieved April 22, 2006 from <http://www.gao.gov/new.items/d06268.pdf>.
- ¹³¹ Ingersoll, R. M. (1999). The problem of underqualified teachers in American secondary schools. *Educational Researcher*, 28(2), 26–37.
- ¹³² Weiler, R.M., Pigg R.M., Jr., & McDermott , R.J. (2003, January). Evaluation of the Florida Coordinated School Health Program Pilot Schools Project. *Journal of School Health* (73)1. pp. 3-9.
- ¹³³ Greenberg, J., Cottrell, R., Bernard, A. (2001). Baseline data on coordinated school health programs in the state of Ohio. *American Journal of Health Studies*, (17)1.
- ¹³⁴ Lubienski, S., & Lubienski, C. (2005). A new look at public and private schools: Student background and mathematics achievement. *Phi Delta Kappan*, 86(9), 696. Retrieved May 15, 2005, from http://www.pdkintl.org/kappan/k_v86/k0505lub.htm.
- ¹³⁵ Braun, H., Jenkins, F., & Grigg, W. (2006). Comparing Private Schools and Public Schools Using Hierarchical Linear Modeling. Washington, DC: National Center for Education Statistics. Retrieved August 3, 2006 from <http://nces.ed.gov/NAEP/pdf/studies/2006461.pdf>.
- ¹³⁶ Greene, J. (2001). An evaluation of the Florida A-Plus accountability and school choice program. New York: Manhattan Institute. Available: http://www.manhattan-institute.org/cr_aplus.pdf.
- ¹³⁷ Greene, J., Howell, W., & Peterson, P. (1997). Lessons from the Cleveland Scholarship Program. Cambridge, MA: Program on Education Policy and Governance. Available: <http://www.ksg.harvard.edu/pepg/pdf/lessclev.pdf>.
- ¹³⁸ Greene, J., Peterson, P., & Du, J. (1997). Effectiveness of school choice: The Milwaukee experiment. Cambridge, MA: Program on Education Policy and Governance. Available: <http://www.ksg.harvard.edu/pepg/other/mil.htm>.
- ¹³⁹ Mayer, D., Peterson, P., Myers, D., Tuttle, C.C., & Howell, W. (2002). School choice in New York City after three years: An evaluation of the School Choice Scholarships Program, final report. New York: Mathematica Policy Research, Inc. Available: <http://www.mathematica-mpr.com/PDFs/nycfull.pdf>.
- ¹⁴⁰ Krueger, A. B., & Zhu, P. (2003). *Another look at the New York City school voucher experiment*. New York: National Center for the Study of Privatization in Education. Retrieved April 22, 2006 from http://www.wvs.princeton.edu/policybriefs/krueger_voucher.pdf.
- ¹⁴¹ Carnoy, M. (1997). Is privatization through education vouchers really the answer?: A comment on West. *The World Bank Research Observer*, 12(1). pp. 105–16. Retrieved April 21, 2006 from <http://www.worldbank.org/research/journals/wbro/obsfeb97/pdf/artcle~7.pdf>
- ¹⁴² Belfield, C.R. (2006, January). *The Evidence on Education Vouchers: An Application to the Cleveland Scholarship and Tutoring Program*. New York, NY: National Center for the Study of Privatization in Education at Teachers College.

- ¹⁴³ Wylie, C. (1998). Can vouchers deliver better education? A review of the literature, with special reference to New Zealand. New Zealand Council For Educational Research. Retrieved April 21, 2006 from <http://www.nzcer.org.nz/pdfs/5835.pdf>.
- ¹⁴⁴ Witte, J. F. (1998, Winter). The Milwaukee voucher experiment. *Educational Evaluation and Policy Analysis*, 20, 229–251.
- ¹⁴⁵ Plucker, J., Muller, P., Hansen, J., Ravert, R., & Makel, M. (2006, February 22). Evaluation of the Cleveland Scholarship and Tutoring Program: Summary Report 1998 – 2004. Bloomington, IN: Center for Evaluation and Education Policy. Retrieved April 22, 2006 from http://ceep.indiana.edu/projects/PDF/200602_Clev_Summary.pdf.
- ¹⁴⁶ Myers, D. & Pfliegerer, J. (2000, September 15). Voucher claims of success are premature in New York City: Second-year results show no overall differences in test scores between those who were offered vouchers and those who were not. *Press release*. Princeton, NJ: Mathematica Research, Inc. Retrieved March 28, 2006 from <http://www.mathematica-mpr.com/Press%20Releases/Past%20Releases/voucherrelfinal.asp>.
- ¹⁴⁷ Krueger, A. B., & Zhu, P. (2003). *Another look at the New York City school voucher experiment*. New York: National Center for the Study of Privatization in Education. Retrieved April 22, 2006 from http://www.wvs.princeton.edu/policybriefs/krueger_voucher.pdf.
- ¹⁴⁸ Camilli, G., & Bulkely, K. (2001, March 4). Critique of “An evaluation of the Florida A-Plus accountability and school choice program.” *Education Policy Analysis Archives*, 9(7). Retrieved April 12, 2006 from <http://epaa.asu.edu/epaa/v9n7>.
- ¹⁴⁹ Kupermintz, H. (2001). The effects of vouchers on school improvement: Another look at the Florida data. *Education Policy Analysis Archives*, 9(8). Retrieved March 21, 2006 from <http://epaa.asu.edu/epaa/v9n8>.
- ¹⁵⁰ Greene, J. (2001). An evaluation of the Florida A-Plus accountability and school choice program. New York: Manhattan Institute.
- ¹⁵¹ Witte, J. F. (1995). Reply to Greene, Peterson and Du: “The effectiveness of school choice in Milwaukee: A secondary analysis of data from the program’s evaluation.” University of Wisconsin, Madison: Robert M. LaFollette Institute of Public Affairs. Available: <http://dpls.dacc.wisc.edu/choice/replytoc.html>.
- ¹⁵² Metcalf, K (1998, September 23). Advocacy in the guise of science. *Education Week*, 18(3), 34, 39.
- ¹⁵³ Hoxby, C.M. (2004). A straightforward comparison of charter schools and regular public schools in the United States. *National Bureau Of Economic Research*. Retrieved April 19, 2006 from <http://post.economics.harvard.edu/faculty/hoxby/papers/hoxbyallcharters.pdf>.
- ¹⁵⁴ Nelson, F.H., Rosenberg, B., & Van Meter, N. (2004, August) Charter school achievement on the 2003 National Assessment of Educational Progress. Washington, DC: American Federation of Teachers. Retrieved April 22, 2006 from <http://www.aft.org/pubs-reports/downloads/teachers/NAEPCharterSchoolReport.pdf>.
- ¹⁵⁵ Miron, G. & Nelson, C. (2001). Student academic achievement in charter schools: What we know and why we know so little. New York, NY: National Center for the Study of Privatization in Education at Teachers College. Retrieved March 23, 2006 from http://www.ncspe.org/publications_files/590_OP41.pdf.
- ¹⁵⁶ Finnigan, K., Adelman, N., Anderson, L., Cotton, L., Donnelly, M. B. & Price, T. (2004). Evaluation of the Public Charter Schools Program: Final Report. Prepared for: U.S. Department of Education, Office of the Deputy Secretary, Policy and Program Studies Service. Washington, D.C.: SRI International. Retrieved June 21, 2006 from <http://www.ed.gov/rschstat/eval/choice/pcsp-final/finalreport.pdf>.

- ¹⁵⁷ 9 Zimmer, Ron & Buddin, Richard (2006) Making sense of charter schools: Evidence from California. Santa Monica, CA: RAND Corp. Retrieved June 18, 2006, from http://www.rand.org/pubs/occasional_papers/2006/RAND_OP157.pdf.
- ¹⁵⁸ Gill, B., Hamilton, Lockwood, J.R., L., Zimmer, R., Marsh., J, Hill, D., Pribesh, S.P.. (2005). *Inspiration, Perspiration, and Time: Operations and Achievement in Edison Schools*. RAND Corporation. Santa Monica: CA.
- ¹⁵⁹ Bush, G. (2002, November 5). *Statement by the President*. Washington, DC: Office of the Press Secretary. Retrieved May 21, 2003 from <http://www.whitehouse.gov/news/releases/2002/11/20021105-4.html>.
- ¹⁶⁰ School Law News. (2002, December). *Bush reserves right to intervene in ED research*. Washington, DC: Aspen Publishers, Inc.
- ¹⁶¹ Gamboa, A.H. (2005, September 30). *Department of Education—Contract to Obtain Services of Armstrong Williams*. Washington, DC: United States Government Accountability Office. Report B-305368. Retrieved April 22, 2006 from <http://www.gao.gov/decisions/appro/305368.pdf>.
- ¹⁶² Gamboa, A.H. (2005, September 30). *Department of Education—No Child Left Behind Act Video News Release and Media Analysis*. Washington, DC: United States Government Accountability Office. Report B-304228. Retrieved April 22, 2006 from <http://www.gao.gov/decisions/appro/304228.pdf>.
- ¹⁶³ Kepplinger, G.L. (2005, September 30). *Department of Education—No Child Left Behind Newspaper Article*. Washington, DC: United States Government Accountability Office. Report B- 306349. Retrieved April 22, 2006 from <http://www.gao.gov/decisions/appro/306349.pdf>.
- ¹⁶⁴ Shaul, M.S. (2006, February). Discretionary grants: Further Tightening of Education’s Procedures for Making Awards Could Improve Transparency and Accountability. *Report to the Ranking Minority Member, Education and the Workforce Committee, House of Representatives*. Washington, DC: United States Government Accountability Office. Report GAO-06-268. Retrieved April 22, 2006 from <http://www.gao.gov/new.items/d06268.pdf>.
- ¹⁶⁵ Manzo, K.K. (2005, October 12). GAO to probe Federal plan for reading: Senate education leaders request an investigation. *Education Week* 25(7). pp. 1/22.
- ¹⁶⁶ Blair, J. (2002, August 7). Critics claim missteps on execution of Title II. *Education Week* 21(43). Retrieved May 12, 2005 from <http://www.edweek.org/ew/articles/2002/08/07/43hea.h21.html>.
- ¹⁶⁷ Earley, P. (2003). *Analysis of the Secretary of Education’s 2003 Report on Teacher Quality: It’s déjà vu all over again*. Fairfax, VA: Center for Education Policy, George Mason University. Retrieved May 12, 2005 from <http://gse.gmu.edu/centers/edpolicy/documents/dejavu.pdf>.
- ¹⁶⁸ Keller, B & Galley, M. (2002, June 19). Paige uses report as a rallying cry to fix teacher ed. *Education Week* 21(41). Retrieved May 10, 2005 from <http://www.edweek.org/ew/articles/2002/06/19/41title2.h21.html>.
- ¹⁶⁹ Carter, G. (2002). ASCD raises concerns about the U.S. Department of Education report on teacher quality. *ASCD Position Statement*. Alexandria, VA: Association for Supervision and Curriculum Development. Retrieved May 12, 2005 from <http://www.ascd.org/portal/site/ascd/menuitem.66824d614f672d9cbfb3ffdb62108a0c/template.article?articleMgmtId=ba173f4062520010VgnVCM1000003d01a8c0RCRD>.
- ¹⁷⁰ Blair, J. (2002, August 7). Critics claim missteps on execution of Title II. *Education Week* 21(43). Retrieved May 12, 2005 from <http://www.edweek.org/ew/articles/2002/08/07/43hea.h21.html>.
- ¹⁷¹ Nelson, F.H., Rosenberg, B., & Van Meter, N. (2004, August) Charter school achievement on the 2003 National Assessment of Educational Progress. Washington, DC: American Federation of Teachers. Retrieved April 22, 2006 from <http://www.aft.org/pubs-reports/downloads/teachers/NAEPCharterSchoolReport.pdf>.

- ¹⁷² National Association for Bilingual Education. (2005, August 30). Good news about bilingual education: Too hot for Feds to handle? Press Release. Washington, DC: Author. Retrieved April 23, 2006 from <http://www.nabe.org/press/press9.html>.
- ¹⁷³ Laitsch D., Heilman, E., & Shaker, P. (2002, December). Teacher Education, Pro-Market Policy and Advocacy Research. *Teaching Education* 13(3). pp. 251-271.
- ¹⁷⁴ Rethinking Schools Online. (2001, Fall). Who's Bankrolling Vouchers? *Special Voucher Report*. Milwaukee, WI: Author. Retrieved March 28, 2006 from http://www.rethinkingschools.org/special_reports/voucher_report/v_bank.shtml.
- ¹⁷⁵ Callahan, D. (1999). *\$1 Billion for Ideas: Conservative think tanks in the 1990s*. Washington, DC: National Committee for Responsive Philanthropy.
- ¹⁷⁶ State Policy Network. (2006). About SPN. Richmond, CA: Author. Retrieved April 22, 2006 from <http://www.spn.org/about/>.
- ¹⁷⁷ Laitsch D., Heilman, E., & Shaker, P. (2002, December). Teacher Education, Pro-Market Policy and Advocacy Research. *Teaching Education* 13(3). pp. 251-271.