

## Annotated Bibliography

---

### **Continuity and Accountability in a Changing Political Environment: Ramifications for Educational Policy Making**

*Prepared by  
Marcelo Gonzalez*

**Wirt, F. M., & Kirst, M. W. (2001). *The political dynamics of American education, 2nd ed.* Berkeley, CA: McCutchan.**

This widely used, well-researched book begins with “an overview of the origins, nature and impact of political forces...influencing schools” (p. x) with discussion of the various issues, actors, and trends in the contemporary educational landscape. In addition, the authors examine the complex, interrelated policy making role of local, state, and federal authorities (including that of the judiciary) and the way in which these entities respond to the diverse demands of a heterogeneous public while addressing the often competing values of equity, quality, and efficiency. The importance of this book to the field lies in its comprehensive analysis of political forces affecting educational policy in the U.S. from an historical and theoretical perspective.

**Hanuschek, E. A., & Raymond, M. E. (2001). *The confusing world of educational accountability. National Tax Journal, 54(2), 365-384.***

In their analysis of the design, impact, and use of accountability systems, the authors begin with a description of the U.S. educational environment. Despite increased resources (as seen in lower pupil-teacher ratios, more highly qualified teachers and higher per-pupil expenditures), U.S. students continued to perform in the 1990s at levels similar to those achieved in the 1960s. The authors contend that lower student performance has resulted in a shift in policy focus from the “traditional regulation of inputs,” (i.e., resource expenditures) to that of outcomes. The authors see this shift reflecting a more general policy trend in the area of regulatory reform and leading to our current “standards based reform.” Finally, the authors carefully discuss the issues of standards, measurements, and consequences (“the three legs” of accountability systems) and their relationship to performance, focusing on the ways in which the various actors of the system are influenced. The article’s importance to the field lies in its comprehensive treatment of the issues involved in educational accountability, which are likely to grow more complex as our society diversifies.

**Herrington, C. D. (2001). *Rethinking the role of the state: Accountability and large-scale systemic reform.* In C. D. Herrington and K. Kasten (Eds.), *Florida 2001: Educational Policy Alternatives* (pp. 227-231), Jacksonville: University of North Florida, Florida Institute of Education.**

In this final chapter of *Florida 2001: Educational Policy Alternatives*, Herrington notes that despite more than 30 years of educational reform, the state “continues to wrestle with the same issues of how to ensure quality, maintain equity, and stimulate improvement.” In summarizing the book’s chapters, the author suggests that the state’s role of accountability must extend beyond oversight and assessing effectiveness to enhancing the capacity of schools and school districts to dramatically improve student performance.

Citing the lack of clear evidence of “what works,” Herrington calls for a broad, large-scale research agenda, providing empirical data from diverse sources and research models to inform policy decisions, and state allocation of resources.

**Ladd, H. F. (2001). School-based educational accountability systems: The promise and the pitfalls. *National Tax Journal*, 54(2), 385-400.**

Ladd (2001) addresses various approaches to educational accountability and discusses her preference for site-based accountability systems, i.e., those that focus on the school (rather than on the district) as their “primary unit” of analysis. Ladd explains the advantages of value-added approaches in measuring school performance, while calling attention to the disadvantages of other methods, which may favor some groups over others. In addition, the author addresses issues in determining the degree to which accountability systems may lead to increased student achievement. Despite (early) indications from Texas and North Carolina, further research is required to ensure reliable interpretation of seemingly positive results. Finally, the author discusses potential implications of the various ways in which site-based accountability systems may influence the behavior of school principals. The article’s importance to the field lies in its (anecdotally supported) analysis of how site-based accountability systems and their concomitant incentives affect the ways in which schools operate.

**Hess, F. M. & Leal, D. L. (2001). Quality, race, and the urban education marketplace. *Urban Affairs Review*, 37 (2), 249-266.**

The findings of this study point to the value of a “market reform” approach to education, in which increased competition leads to improved school performance. Based on data from 50 urban areas (with populations of 50,000 or more), the authors found that community perceptions of school quality influence private school enrollment, as do community perceptions regarding racial and religious compositions of public schools, although to a lesser extent. Further research is needed to determine the relative influence of the identified factors. The authors note that this is the first quantitative study of factors influencing school enrollment that focuses on large urban areas, the population centers most likely to benefit from school choice programs.

**Saporito, S., Yancey, W. L., & Louis, V. (2001). Comments: Quality, race, and the urban education marketplace reconsidered. *Urban Affairs Review*, 37 (2), 267-276.**

In their response to Hess & Leal (2001), the authors take issue with both the study’s findings and its underlying assumptions, particularly the portrayal of parents as informed consumers in an educational marketplace who make enrollment decisions based on the instructional quality of schools rather than on the socioeconomic status and racial composition of their student bodies. Using data from various sources, including those of

Hess & Leal (2001), Saporito and his colleagues take issue with their counterparts' assertion that higher private school enrollments are not likely to cause lower public school graduation rates. They found a clear relationship between higher socioeconomic status and both increased private school enrollments and graduation rates (the latter being the measure of "school quality" employed in Hess & Leal, 2001) and argue that this relationship results in "creaming effects" (i.e., wealthy students attending private schools and the concomitant negative effects on public schools' performance). This study questions the degree to which public school quality may influence parental decisions in choosing schools, and provides evidence of the potential role of "non-rational" considerations (e.g., the desire to distance one's children from those perceived to be of a lower social status) in the decision-making process.

**Reed, D. S. (2001). Not in my schoolyard: Localism and public opposition to funding schools equally. *Social Science Quarterly*, 82, 34-50.**

In his study focusing on Connecticut and New Jersey, the author examines the varying influence of a wide range of demographic and attitudinal factors on opposition to state-imposed equality in public school funding: age, income, race, education, party affiliation, ideology, and parenthood as well as survey data on "taxation," "economic self-interest" (i.e., "not from my taxes"), and "attitudes toward public schools." The study's findings indicate that while a majority of survey respondents in both states support equality in educational funding, race-based differences exist, as "...nonwhites in both states favor equality at substantially higher rates than whites" (p. 42). The most significant finding is the influence of localism. Survey respondents were less likely to support equality of educational funding when such equalization reduces local control. The importance of this study lies in its empirically supported finding that issues of local control may be stronger predictors of opposition to states' plans to equalize public school funding than demographic and other attitudinal factors. This finding may be especially true in the 18 states (including Connecticut and New Jersey) where such reforms have been mandated by state supreme courts (p. 35).

---

### **Closing Florida's Achievement Gaps**

*Prepared by  
Alissa R. Peltzman  
and Madhabi  
Chatterji*

**Buttram, J. L., & Waters, J. T. (1997). Improving America's schools through standards-based education. *NASSP Bulletin*, 81 (590), 1-6.**

The authors offer a timeline of the standards-based movement that begins with the publication of the Coleman Report in 1966 and is redefined by the release of *A Nation at Risk* in 1983, marking a distinct switch in emphasis from inputs to output results and accountability. The authors explain the link between aligning state standards and accountability, taking note of both the number of states setting specific standards in core

subjects as well as the work of the National Council of Teachers of Mathematics (NCTM) and the American Association for the Advancement of Science (AAAS). The authors explicitly support standards-based education, arguing that it delineates specific expectations for student learning, concentrates on instruction of content, and can potentially inform curriculum resource allocation as well as professional development programs for teachers and school leaders. Finally, the authors warn of potential misuse of performance and content standards, which may occur when standards are set too low or too high, without adequate allocation of resources, alignment of assessment, or time for implementation by schools and teachers.

**Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, 20 (4), 19-27.**

The author presents ten positive outcomes of high-stakes standardized testing. The author provides a historical context in which to understand the current negative backlash against high-stakes testing. In the author's view, the main criticism focuses on the consequences of "teaching to the test." The author presents an overview of evidence supporting the negative consequences of high-stakes testing, contending that it is meager. The author endorses the notion that external high-stakes tests should not be the only measure used to make decisions about students; rather, he argues that decisions, such as those regarding promotion or graduation, should be based on multiple criteria and reliable information. The author links resistance to high-stakes testing within the education community to the status quo, in which accountability is not centered on teacher performance. The author claims high-stakes testing has prompted much-needed improvements in K-12 professional development. The author's list of benefits includes attempts to extend testing accommodation for students with special needs, creating more inclusive classrooms; the inclination for educators to learn about testing and dispel previously held myths about assessment; and an improvement in the quality, reliability, objectivity, and relevance of tests themselves.

**Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational Researcher*, 31(1), 3-12.**

The author analyzes trends and contributing factors regarding the Black-White and Hispanic-White achievement gap through secondary analysis of national data sets. The author utilizes National Assessment of Educational Progress (NAEP) data to demonstrate that U.S. students made moderate gains in reading scores from 1971-1999, and in mathematics scores from 1973-1999, but notes that this improvement masks the disparities between students of different ethnic groups. The author uses time-series regression analyses to establish that the Black-White achievement gap narrowed until the mid-1980s and then increased as standards-based reforms took hold (Blacks improved,

while Whites remained stagnant through the 1980s, while Whites improved and Blacks slid lower in the 1990s). The Hispanic-White gap had a similar pattern of improvement and decline. The author concludes that this trend corresponds to the shift from minimum competency testing to the standards movement. He states that minority student gains were greater when there was an emphasis on basic skills, whereas White students performed better when higher standards were emphasized. The author attempts to reveal discrepancies and demonstrate the extent to which instruction had a disparate impact on different ethnic groups. The author also examines family dynamics, socio-economic factors, peer interaction, and changes in schooling as factors affecting achievement differences in ethnic groups.

**Linn, R. L. (2003). *Accountability: Responsibility and reasonable expectations. Educational Researcher, 32(7), 3-13.***

The author explains broadly that shared responsibility among all stakeholders is necessary for an accountability system to improve public education. The author cites the advantages of using National Assessment of Educational Progress (NAEP) achievement levels as a point of reference for measuring the strength of state standards and assessments, and to determine how realistic expectations are for progress and long-term goals of schools. The author recommends that there should be a healthy relationship between research and policy, in this case regarding the design and implementation of accountability systems, that will lead to improved schools. In particular, the author is concerned that the ambitious goals for Adequate Yearly Progress under the No Child Left Behind Act are too high and therefore out of reach for most schools.

**Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). *Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. Educational Researcher, 31(6), 3-16.***

The authors praise the ambitious goals of *No Child Left Behind Act* of 2001 (NCLB) but point to implementation challenges, particularly in regards to variance in current state standards and assessments. They show that different states have vastly different starting points, and demonstrate the challenge for schools in posting significant gains when states utilize rigorous assessments with high-order thinking. The authors use the National Assessment of Educational Progress (NAEP) data to demonstrate rates of school improvement that are feasible, and in doing so, argue that the objectives for Adequate Yearly Progress as stipulated in the NCLB are unrealistic. The authors detail challenges posed by variability in test scores at the individual school level each year. They recommend expanding the accountability criteria, specifically endorsing use of an index score and considering using NAEP instead of individual state tests as a means of providing more reliable and equitable data.

**Morris, D. R. (2001). Assessing the implementation of high-stakes reform: Aggregate relationships between retention rates and test results. *NASSP Bulletin*, 85 (629), 18-34.**

The author supplies an analysis of retention as it relates to high-stakes test scores in Florida. The author analyzes the relationship between high-stakes test data and grade retention in Miami-Dade County, controlling for socio-economic status between 1986 and 1991. With the exception of 7th grade, the data revealed a positive relationship between grade retention rates and the percentage of students scoring below the 50th percentile. That is, the study found that the higher the retention rate in a school, the higher was the percent of students scoring below the median on the nationally normed test. The author also demonstrates that there was a curvilinear relationship and submits that the retention policy was in fact effective in affluent schools, although not with high-risk student populations. The author points out that retention associated with the basic skills testing movement in the 1980s was discontinued by the state legislature because it failed to raise student achievement standards. The author also challenges current student remediation policies that use grade retention practices in Florida, due to factors such as formidable costs, lack of classroom space, and pressure for test information to get distorted.

**Ogbu, J. U. (1994). Understanding cultural diversity and learning. *Journal for the Education of the Gifted*, 17 (4), 355-83.**

Ogbu addresses underachievement and academic disengagement of African American students in suburban high schools, where poverty is not a factor. He suggests that African American students learn and perform well if teachers “pour knowledge” into them, that African American parents do not perceive themselves as active agents in the education process, and that African Americans hold teachers accountable for students’ performance. This article argues that neither the core curriculum movement nor the multicultural education movement adequately addresses the problem of minority groups that do not perform well in public schools. Minorities whose cultural frames of reference are oppositional to the American mainstream culture have difficulty crossing cultural boundaries to learn in school.

**Ogbu, J. U. (1999). Ebonics, proper English, and identity in a black American speech community. *American Educational Research Journal*, 36 (2), 147-84.**

Ogbu describes and explains the socio-linguistic factors that affect performance of African Americans in speaking standard English. He uses data from a two-year study of 40 adults and 76 students to show how the black community has difficulty learning standard English due to incompatible beliefs about that dialect.

**Sobol, T. (2003). *The No Child Left Behind Act: Recommendations for what schools and districts should do*. Unpublished speech for school and district leaders: November 10, 2003. New York: Teachers College, Columbia University.** The author illustrates the benefits of the standards movement: providing instructional focus, prompting higher quality instruction, promoting equity and consistency. The author recognizes the potential value of standards, testing, and accountability, noting that success and usefulness depend on the type of tools used to implement such reform. The author argues that the current backlash against this movement is due to inevitable problems as the policy gets underway, implementation challenges, and a limited vision of teaching and learning within the federal *No Child Left Behind Act*. The author maintains that the standards movement has limited curriculum and led to an emphasis on rote learning. The author submits that the policy timeline is not consistent with how schools operate, students are not provided sufficient opportunities to prepare for high-stakes tests, and education policy is shifting from local to more federal control. The author recommends that tracking, promotion, and graduation decisions be made based on multiple criteria, including academic achievement test scores. The author proposes utilizing multiple assessments, reducing the overall amount of testing, increasing funding, and developing local capacity for a sensible and flexible accountability plan that promotes high-quality education.

---

### **Florida's Constitutional Amendment to Reduce Class Size: What Can Research Tell Us?**

***Prepared by  
Doug Harris***

**Berliner, D., & Biddle, B. (2002). *Small class size and its effects*. *Educational Leadership*, 59(5).**

The authors provide a literature review of class size research, going back to the Coleman Report and including the STAR and SAGE experiments (described below). They argue that the pupil-teacher ratio, used in a large number of correlational studies, is a poor proxy for class size. They therefore give greater weight to the positive effects found in SAGE and STAR. The authors discuss a large number of interrelated theories about why class size reduction is effective, including increased teacher morale, increased one-on-one interaction between students and teachers, improved classroom environment, fewer student discipline problems, and greater help for students in coping with traditional classroom settings to which they may be unaccustomed. This last theory, they argue, explains why disadvantaged students appear to make greater achievement gains than other students who attend small classes. The authors also raise the issue of costs, although they dismiss previous studies that have considered the topic. They conclude that "if citizens are truly committed to providing a quality public education and a level playing field for all students regardless of background, they will find the funds needed to reduce class size" (p.22).

**Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Willms, J. D. (2001). Class size and student achievement, *Psychological Science in the Public Interest*, 2(1), 1-30.**

The authors provide an even more extensive literature than Biddle and Berliner (2002), including substantial detail about the methodological issues related to identifying the effects of class size reduction. They argue that experimental studies are best, although the results cannot always be generalized to other settings. One reason is that experiments tend to involve a small number of participants (students and teachers). Correlational studies can be more easily generalized because they have larger and broader samples. On the other hand, the latter studies are lacking because of the greater potential for selection bias and other factors that can be minimized in experiments where participants are randomly assigned. Like other authors, they find that the STAR experiment produced positive effects, although they describe some potential weaknesses of the study. The correlational research they summarize indicates less consistent effects from class size. The authors do not take a hard stand on what are the true benefits of class size. Instead, they argue for additional research using experimental research designs, which examine not only whether class size matters, but how and why.

**Finn and Achilles (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analyses*, 21(2), 97-110.**

The authors provide a literature review of evidence on class size reduction, especially the Tennessee STAR experiment. The experiment began in 1985 and involved 12,000 students who were randomly assigned to small classes (13-17 students) and large classes (22-26 students). The authors of the article support the notion that class size has a positive and lasting effect on student achievement. They explicitly refute the idea that class size matters only in kindergarten, which has been argued by Hanushek based on his interpretation of STAR. They also refute evidence from the large number of correlational studies that have used the pupil-teacher ratio as a proxy for class size, the same point made by Biddle and Berliner. The authors also discuss their own theories about why class size reduction works: that smaller classes ease classroom management and increase student engagement.

**Molnar, A., Smith, P., Zahorik J., & Palmer, A., Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Educational Evaluation and Policy Analyses*, 21(2), 97-110.**

This study provides an analysis of the Wisconsin SAGE program, which differs from the STAR experiment in several ways: (a) students and teachers were not randomly assigned; (b) the program was applied only in high-poverty schools; and (c) the program included after-school programs and other elements. The achievement effects were similar, although

somewhat larger. The authors argue that the entire effect can be attributed to class size because the other elements were implemented more slowly. The results also differed from STAR in several important ways: As in STAR, class sizes were “reduced” in some classrooms by having two teachers with approximately 30 students in the same classroom, in addition to more traditional class size reductions in which one teacher is matched with roughly 15 students. The STAR study found that adding a teacher to a large classroom has no effect. The results with SAGE suggest the opposite: achievement gains were the same in the 30:2 and 15:1 classrooms, suggesting that adding a teacher to a classroom is important. This also leads the authors to present a somewhat different theory about why class size reduction works. Without referring to STAR or the above study, they argue that class size reduction works because of the additional individual attention from teachers, not simply student engagement.

---

### **Building Capacity for Teaching and Learning in Florida**

*Prepared by  
Katherine Kasten  
and Roderick J.  
McDavis*

Several researchers have studied the ability of urban communities to initiate, support, and sustain school reform. Hill, Campbell, and Harvey (2000) studied six cities for two years and also drew on case studies of five other cities. They concluded that reform must be supported by citywide strategies that combine “incentives for school performance, investments in school capabilities, and freedom for change” (p. 63). Another national study of 11 cities considered how cities have mobilized to support school change (Stone, 2001; Stone, Heniz, Jones, & Pierannunzi, 2001). These researchers noted that schools have often been used “to try to reform society on the cheap” (p. 53). They argue, instead, that society should consider how social reform might support and enhance the work of schools. Defining civic capacity as “the extent to which different sectors of the community...act in concern around a matter of community-wide impact” (Stone, p. 596), the researchers identify Atlanta as a low-capacity city, Chicago as having moderate capacity, and El Paso as demonstrating high capacity. Finally, Hess (1999) argued that urban school reform so often fails because rewards accrue to school boards, superintendents, and other policy makers for initiating, rather than implementing, change. Using abbreviated case studies of 57 urban districts, Hess contrasts the short-term effects of sweeping attempts at change with those of controlled, deliberate efforts and the impact of outsiders versus insiders on the change process. Hess concluded that “a state of perpetual reform is the status quo,” a phenomenon he refers to as “policy churn.”

#### **Sources:**

Hess, F.M. (1999). *Spinning wheels: The politics of urban school reform*. Washington, DC: Brookings Institution Press.

Hill, P. T., Campbell, C., & Harvey, J., et al. (2000). *It takes a city: Getting serious about urban school reform*. Washington, DC: Brookings Institution Press.

Stone, C. N. (2001). Civic capacity and urban education. *Urban Affairs Review*, 36, 595-619.

Stone, C. N., Henig, J. R., Jones, B. D., & Pierannunzi, C. (2001). *Building civic capacity: The politics of reforming urban schools*. Lawrence: University of Kansas Press.

Richard Elmore has called attention to the fact that key aspects of implementing standards-based school reforms run counter to the organization and culture of American schools. Elmore argues that increasing academic quality and performance requires investment in capacity building through developing the knowledge and skills of teachers and administrators. In fact, one of Elmore's fundamental propositions is that accountability and capacity-building are reciprocal responsibilities. Elmore argued that reform requires identifying schools (as opposed to school districts or individual teachers and classrooms) as the unit of accountability, distributing leadership in the school, and changing the ways that schools manage resources. Elmore also noted that policy makers rarely have the knowledge and skill to change the core of instruction. To effect instructional change, knowledge of policy and knowledge of teaching and learning practice must be overlaid on each other.

**Sources:**

Elmore, R. F. (2000). *Building a new structure for school leadership*. Washington, DC: Albert Shanker Institute.

Elmore, R. F. (2002). *Bridging the gap between standards and achievement: The imperative for professional development in education*. Washington, DC: Albert Shanker Institute.

Elmore, R. F., & Burney, D. (1999). Investing in teacher learning: Staff development and instructional improvement. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 263-291). San Francisco: Jossey-Bass.

One of the most prolific writers in support of the professionalization of teaching is Linda Darling-Hammond. Darling-Hammond argues that effective teaching is fundamental to educational reform. Drawing on the research of others, she has demonstrated that effective teachers are differentially distributed across American schools and school districts and that poor and minority children are more likely to be taught by under-prepared and unqualified teachers than are white or affluent children. As a strong

advocate of rigorous requirements for teacher credentialing as well as extensive local and state investments in ongoing professional development for teachers and administrators, Darling-Hammond has often been at odds with those who support reducing the requirements for obtaining a teaching license and providing multiple and flexible pathways into teaching.

**Sources:**

Darling-Hammond, L. (2000). Teachers for America's future: National commissions and vested interests in an almost profession. *Educational Policy*, 14, 162-183.

Darling-Hammond, L. (1999). *Teacher quality and student achievement: A review of state policy evidence*. Seattle: Center for the Study of Teaching and Policy, University of Washington.

Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco: Jossey-Bass.

Darling-Hammond, L., Berry, B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis*, 23, 57-77.

Darling-Hammond, L., & Sykes, G. (Eds.). (1999). *Teaching as the learning profession: Handbook of policy and practice*. San Francisco: Jossey-Bass.

David Cohen and Deborah Ball have written extensively together and with others about the development of instructional capacity in schools and teachers. They note several impediments to significant improvement of teaching and learning: lack of depth and sophistication in teachers' knowledge of content, lack of student engagement, lack of coherent guidelines for instruction, and lack of continuing opportunities for teachers to improve their practice. They argue that increasing teachers' opportunities to learn is a crucial feature of improving instructional capacity, and capacity can be developed through teachers becoming serious students of their own teaching practice. Defining capacity as a combination of resources, knowledge, and support, Ball and Cohen argue that capacity is not a fixed attribute but a "feature of interaction" among these elements. In a recent (2003) article, Cohen and Ball define numerous supports for improved teaching and learning: conventional resources such as materials, facilities, class size, and time; personal resources such as teachers' will, skill, and knowledge; and contextual resources such as academic norms, political and professional leadership, and family support. Noting that research on features such as reduced class size and increased instructional time has only modestly supported a connection to increased student achievement, they argue that resources are only valuable as teachers and students know

how to use them and parents and community support their use.

**Sources:**

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3-32). San Francisco: Jossey-Bass.

Cohen, D. K., & Ball, D. L. (1999). *Instruction, capacity, and improvement*. CPRE Research Report Series RR-43. Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.

Cohen, D. K., Raudenbush, S. W., & Ball, D. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*, 119-142.

A number of studies are now available that describe what standards-based reform looks like at the district, school, and classroom levels. Much of this work comes from standards-based reform efforts in Kentucky, Tennessee, Maryland, and other states; reform sponsored by the National Science Foundation; and the comprehensive school reform initiatives identified in Title I legislation. The sources listed are only a few examples.

**Sources:**

Borko, H., Wolf, S. A., Simone, G., & Uchiyama, K. P. (2003). Schools in transition: Reform efforts and school capacity in Washington State. *Educational Evaluation and Policy Analysis, 25*, 171-201. Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research, 72*, 433-479.

Spillane, J. P. (2001). Challenging instruction for "all students": Policy, practitioners, and practice. In S. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states* (pp. 217-241). 100th Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press.

Spillane, J. P., & Thompson, C. L. (1997). Reconstructing conceptions of local capacity: the local education agency's capacity for ambitious instructional reform. *Educational Evaluation and Policy Analysis, 19*, 185-203.

Wilson, S. H., & Floden, R. E. (2001). Hedging bets: Standards-based reform in classrooms. In S. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states* (pp. 193-216). 100th Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press.

Wolf, S.A., Borko, H., Elliott, R. L., & McIver, M. C. (2000). "That dog won't hunt!": Exemplary school change efforts within the Kentucky reform. *American Educational Research Journal, 37*, 349-393.

Several data sources provide state-by-state comparisons. The ones used in preparation of this policy brief included the annual *Quality Counts* report published by *Education Week* (see <http://www.edweek.org>), the annual rankings and estimates data published by the National Education Association (see <http://www.nea.org>), the recently available education rankings published by Morgan Quinto Press as the “smartest state” award (see <http://www.morganquitno.com>). Florida’s grades according to *Education Week* are generally high on accountability (A- in 2002, A in both 2003 and 2004) and resource equity (B in both 2002 and 2003, C+ in 2004) and low on improving teacher quality (D+ in 2002, C- in 2003, C+ in 2004) and resource adequacy (C- in 2002 and D+ in 2003 and 2004). Those evaluations are consistent with the claim in this policy brief that Florida’s investment in accountability has not been matched by an equal investment in the development of educational capacity.

---

### **The Cost of Accountability: Considerations and Assumptions**

*Prepared by  
Yas Nakib*

In accounting for the cost of accountability, one has to consider both direct and indirect costs. Direct costs of accountability, especially when measured per-pupil, are miniscule compared to the total average per pupil cost. But a careful account of all costs must include indirect costs which could very well be relatively high, depending on the elements of accountability (such as the types of tests used) and its administration. Whether the cost of accountability is high or low, and whether in effect it is worthwhile, are questions that can be accurately answered only by the careful comparison of all costs and benefits which are often complex to account for and measure.

There are many good readings on estimating the cost of accountability. Among the most comprehensive frameworks for analyzing the cost of assessment is the work by Picus (1994). He provides a comprehensive list of ingredients that are needed to estimate expenditures for alternative assessment programs and uses a three-dimensional matrix to identify the costs. One dimension is the components of the assessment. The second is the level (classroom, school, district, state, or national) at which expenditures occur, and the third captures the kinds (personnel, materials, supplies, etc.) of items purchased for each component at each level.

An alternative, but similar, framework is provided in Phelps, R (2000). The cost of student testing activities can be measured by adding expenditure “objects” or “functions.” “Objects” include purchased material and services, time spent on the activities by schools, districts, and state agencies, time of students preparing and taking the test, and administrative and building overhead. “Functions” include start-up test development; on-going test development; preparing students; training for administration; preparing the administration; training for scoring; scoring, collecting, sorting and mailing the tests;

analyzing and reporting the scores; and miscellaneous other activities.

The cost of administering testing and assessment can differ with time. Phelps (2000) and the United States General Accounting Office (2003) found that the most difficult test components to measure are personnel costs and the amount of time devoted to test-related activities. Because expenditures for test development are small relative to non-development expenditures (test administration, scoring, and reporting), estimates of expenditures may be lower in the first few years of test development than in later years.

**Sources:**

Picus, L. (1994). *A conceptual framework for analyzing the costs of alternative assessment*. CSE Technical Report 384. Los Angeles: CRESST, University of California-Los Angeles.

Phelps, R (2000). Estimating the cost of standardized student testing in the United States. *Journal of Education Finance*, 24, 343-380.

United States General Accounting Office (2003). *Title-I: Characteristics of tests will influence expenses; Information sharing may help states realize efficiencies*. GAO-03-389. Washington, DC: GAO.

Some of the other complexities of estimating costs are discussed by Figlio (2004), who indicates that the type of students to be tested and any additional accommodation required to take the test can easily drive up the costs. Also, student mobility within and across districts could be a factor contributing to cost increase. Hanushek, Kain, & Rivkin (2001) also demonstrate that there are variations in the quality of resources (such as teachers) in certain schools and school districts, and that would certainly complicate and drive up costs, especially in resource-poor districts.

There are significant indirect costs associated with student assessment. Those indirect costs can be measured by estimating the foregone benefits. Haney, Madaus, & Lyons (1993) estimate the opportunity cost of testing by using net present value of foregone wages as a result of test preparation, and teacher and student time.

**Sources:**

Figlio, D. (2004). Funding and accountability: Some conceptual and technical issues in state aid reform. In J. Yinger (Ed.), *Helping children left behind: State aid and the pursuit of educational equity*. Cambridge, MA: MIT Press.

Hanushek, E., Kain, J., & Rivkin, S. (2001) *Why public schools lose teachers*. NBER Working Paper No. 8599. Cambridge, MA: National Bureau of Economic Research.

Haney, W., Madaus, G., & Lyons, R. (1993). *The fractured marketplace for standardized testing*. Boston: Kleuwer Academic Publishers.

There have been many studies and estimates of the cost of testing in states and across the nation. Some have reported a very low amount relative to the average per-pupil expenditures in states and the nation. A fairly easy and less complex study by Hoxby (2002) analyzed the educational accountability costs of 25 states that had by 2001 implemented a fairly rigorous system. She also compared the cost of accountability to that of class size reduction to show that, relatively, accountability measures cost much less than other current popular school reforms. Similar and useful studies are listed below.

**Sources:**

Hoxby, C. (2002). *The cost of accountability*. NBER Working Paper No. 8855. Cambridge, MA: National Bureau of Economic Research.

United States General Accounting Office (1993). *Student testing: Current extent and expenditures, with cost estimates for a national examination*. GAO-93-8. Washington, DC: GAO.

United States General Accounting Office (2003). *Title-I: Characteristics of tests will influence expenses; Information sharing may help states realize efficiencies*. GAO-03-389. Washington, DC: GAO.

Picus, L., Tralli, A., & Tacheny, S. (1996). *Estimating the costs of student assessment in North Carolina and Kentucky: A state-level analysis*. CSE Technical Report 408. Los Angeles: CRESST, University of California-Los Angeles.

Monk, D. (1994). *The costs of systemic education reform: A summary report*. Ithaca, NY: Department of Education, Cornell University.