

Association for Public Policy Analysis and Management

**2007 Spencer Foundation Lecture
in Education Policy and Management**

“Holding Schools Accountable Revisited”

By Helen F. Ladd¹

© 2007 Association for Public Policy Analysis and Management.
All rights reserved.

¹ Edgar Thompson Professor of Public Policy Studies and Professor of Economics, Sanford Institute of Public Policy, Duke University. This paper is a somewhat expanded version of the Spencer Foundation Lecture in Education Policy and Management presented by the author at the 2007 APPAM Fall Research Conference in Washington D.C. on November 8, 2007. The author is grateful to the Spencer Foundation and APPAM for selecting her to give this lecture. She also thanks the Spencer Foundation for funding much of her recent research, and her many coauthors who have enriched her research in education policy and management over the years.

“Holding Schools Accountable Revisited”

By Helen F. Ladd

Two years ago, in the first lecture in this series, Eric Hanushek asserted that policy research has improved greatly over time (Hanushek, 2005). I agree with that statement. He then posed the question: If our research has gotten so much better, why hasn’t policy improved? The answer he offered was our failure as researchers to be directly involved in the policy making process.

I agree with him that the more we as researchers can interact with policy makers, the more relevant and useful our research and advice will be – which in turn could potentially generate better policies, though not necessarily in the short run. Even at its best, however, research is only part of the policy making process. Most policy making involves competing values that ultimately must be reconciled through the political process. The challenge for policy researchers is to contribute to value-laden policy debates in ways that maintain our integrity and objectivity while not ignoring the importance of values.

I welcome the opportunity to reflect on my own research on one topic in the value-laden field of education policy – namely policies designed to hold schools accountable for student outcomes.

By school based accountability programs, I am referring to systems that use measures of student outcomes – primarily student achievement as measured by test scores – to hold schools accountable for improving the performance of their students. The federal No Child Left Behind Act (NCLB) of 2001 is obviously the most prominent example. That legislation requires every state to test all students in reading and math annually in grades 3-8 and once in high school. It uses those test scores, reported separately by racial and income subgroups within schools, to hold individual schools accountable for making adequate yearly progress toward the ultimate goal of 100 percent proficiency. Many states, particularly southern states such as my own state of NC, had their own quite sophisticated accountability systems well before the federal law spread school based accountability to all states.

This type of top-down administrative system differs from other forms of accountability, such as political accountability which would hold policy makers accountable through the

political process, or market based accountability in which schools are directly accountable to parents. There is much to say about these other approaches, but I will restrict myself here to administrative accountability.

This is a natural topic for me in part because my first book on education policy was entitled, *Holding School Accountable: Performance Based Reform in Education* (Ladd, 1996) That book emerged from a conference I organized at the Brookings Institution during the mid 1990s at a time when the concept of holding schools accountable for student achievement was relatively new. Since then, I have written a number of papers on related topics and I am currently a member of the North Carolina Blue Ribbon Commission on Testing and Accountability. Hence, the time is ripe for me to step back and reflect on where we have been and where we might go from here.

In addition, of course, the topic is timely in light of the ongoing congressional discussions about NCLB and the reauthorization of the Elementary and Secondary Education Act. I hasten to add that I will have less to say about many of the specific issues under discussion than about some of the broader issues that we, as policy researchers, have the luxury of reflecting upon. . In this paper, I address the following three questions:

To what problem is school accountability the proposed policy solution?

What does the research show?

What is wrong with the current approach and where should we go from here?

The theme throughout is the complex relationship between values, policy and research, a theme to which I will return in the conclusion.

To What Problem Is School Accountability the Proposed Policy Solution?

There are at least three answers to this question, each of which rests on a different set of values. For the proponents of standards based reform, the perceived problem is that the fragmented and incoherent nature of our education system is not well designed to promote the ambitious educational outcomes required in this increasingly global society. For a second group, the perceived problem is that teachers and schools are shirking because it is difficult for the public to monitor their performance. For a third group, the problem is the huge disparities in educational outcomes across groups defined by race or by income.

I initially came to accountability through the first perspective, namely the standards-based reform movement and the work of Mike Smith, Susan Fuhrman and other researchers connected with the Consortium on Policy Research in Education (CPRE). The key values underlying that perspective are high student achievement for all students through systemic reform. Standards-based reform involves the setting of ambitious standards in a set of core academic subjects, establishing high expectations for all students, aligning instruction and professional development to the standards, and assuring all students have an opportunity to learn to the standards (Smith and O’Day, 1991). Testing students is a key component of this reform strategy so that policy makers can measure student progress toward mastery of the standards.

As part of this overall reform strategy, the students’ test scores then serve as the basis for judging the effectiveness of individual schools and for establishing a system of rewards for effective schools and sanctions, or, as appropriate, additional support, for ineffective schools. From this perspective, school based accountability is just one part of a larger reform effort designed to promote higher student achievement.

As such, it is not cheap. By identifying schools in which students are failing to meet the standards, the accountability system puts pressure on higher level policy makers, as least in theory, to provide additional capacity or technical expertise to the schools that need it. Thus, the full cost of the accountability system is not just the costs of developing and administering tests and reporting the results, but also the additional spending needed so schools can meet the specified goals. Furthermore the reform effort is intended to be positive and constructive and not punitive.

But school based accountability can also be seen as a stand-alone policy designed to address the perceived problem that educators are shirking their responsibilities and simply are not working hard enough or “smart” enough to generate the desired outcomes. Economists often use the language of the principal agent model to describe this situation. In the context of such a model, the challenge is to set up an appropriate incentive system to induce the agents – in this case the educators – to operate in ways compatible with the interests of the principal – in this case state policy makers and the public.

Underlying this approach is the value of economic efficiency – that is, promoting greater output with no increase in expenditure. By measuring, reporting and attaching positive

consequences to strong performance and negative consequences to weak school performance, policy makers provide incentives for schools and school districts to focus attention of what is being measured, and ultimately to alter the way they operate. Concerns about the capacity of schools to respond or about inadequate resources clearly take a back seat to confidence in the power of incentives and sanctions to change behavior.

When viewed from this second perspective, accountability appears to be an inexpensive policy intervention. The main costs are setting up testing systems and reporting the scores. The test scores then serve as a catalyst to motivate the desired behavior. Importantly, the associated incentives include both carrots and sticks, but the sticks inevitably play a larger role the greater is the failure of the carrots to achieve the desired outcomes.

Finally, as has been emphasized by groups such as the Citizens Commission on Civil Rights and the Education Trust, school accountability – especially as implemented under NCLB – serves as a tool for addressing the problem of educational inequities. By setting high standards for all students and by focusing attention on the students who the education system has been leaving behind, namely minorities, students from low income families and those who are disabled, school accountability programs serve to promote social justice and equity.

Thus, embedded in the three different conceptions of the policy problem to which accountability is the answer are the values of constructive and comprehensive structures to promote higher student achievement, efficient use of educational inputs, and equity as defined by high educational outcomes. Though the three conceptions are not mutually exclusive, some people clearly put more emphasis on one than another.

The convergence of these three perspectives into support for a common policy – namely, school based accountability, with separate accounting within schools by racial and income subgroup – may help explain the political viability, at least to date, of the accountability movement. The potential conflicts among the values, however, also helps to explain the intensity of the debate and lack of clarity about where we should go from here.

Lest you have any doubt about my values and where I am coming from, you should know that I am far more sympathetic to the equity agenda, and to reforms that are positive and constructive than to those that rely heavily on punitive sanctions.

What Does the Research Show?

Regardless of the values that underlie support for administrative accountability systems, research on such systems can be extremely useful for identifying their effects and highlighting additional policy and value tradeoffs. I turn now to the findings of my research and also that of others on three interrelated topics.

1. Accountability systems can be powerful tools for changing people’s behavior, but not always in intended directions.

First, research shows that accountability systems can be powerful tools for changing the behavior of educators. This result is positive if the changes are in the desired direction but not great if they go in unintended and undesired directions. Moreover, reasonable people can disagree on whether the net effects are positive or negative.

That accountability systems can be powerful tools for changing behavior emerged clearly from some systematic surveying that a colleague and I did of elementary school principals in the first and third years of North Carolina’s highly touted accountability system (Ladd and Zelli 2002). We found, for example, that, consistent with the state’s goal of focusing attention on the basics skills of reading and math, many principals redirected resources to those subjects, increased their work with teachers to prepare for the end-of-grade tests and to improve instruction, and incorporated math and reading into other course and extra-curricular activities.

While many of these actions were consistent with state goals, they also provide support for a well known theorem in the principal-agent literature about the effects of incentive programs in organizations with multiple goals (Milgrom and Roberts, 1992). When only some of the goals can be measured and rewarded, people will focus most of their attention on the rewarded goals to the detriment of the other goals (Gibbons, 1998 and Milgrom and Roberts, 1992, pp. 228-31). Whether that is good or bad in this case involves value judgments. While some people believe the increased attention to the basic skills of math and reading offsets any reduced attention to other subjects, others would disagree and fault the approach for narrowing the curriculum.

Additional evidence of how accountability systems alter behavior emerges from studies of how school officials have gamed or manipulated the system to improve their students’ test results. Several studies have shown, for example, how schools have selectively assigned low

achieving students to special education programs to keep them out of the testing pool (Cullen & Reback, 2006; Deere & Strayer, 2001; Figlio and Getzler, 2006; Jacob, 2005). Further as David Figlio and various coauthors have documented, schools have also changed their meal programs around test time to increase achievement and have engaged in selective disciplinary policies toward the same end (Figlio, 2006 and Figlio and Winicki, 2005).

One lesson I draw here is that accountability systems must be used with care. I like the image introduced in a 1992 paper by Bryk and Hermanson of education being like a rich and complex tapestry. The danger with an accountability system, they argued, is that by pulling on only the most visible strands of the tapestry, the result will be “stress, strains and distortions” (Bryk and Hermanson, 1992, p. 463). One could keep these distortions to a minimum by limiting the magnitudes of any rewards or sanctions so as to minimize the incentives to game the system or to alter people’s behavior in unintended directions. Of course, the flipside is that the positive effects may be mitigated as well so ultimately it is a matter of getting the right balance – with the right balance differing depending on one’s values.

Another lesson that emerges from studies of the effects of accountability takes me to my next point, namely, that the design of the system matters.

2. Design matters – status vs. growth

Among the many issues that arise in the design of accountability systems, I have time to focus on only one, but it is a central design issue in current policy debates about NCLB. This is the question of whether to use a status model or a model based on individual student growth to judge the effectiveness of individual schools. A status model essentially looks at levels of achievement – typically defined as the percent of students who reach a designated level of proficiency – while a growth model – often called a value-added model – focuses on the average gains in learning of individual students from one year to the next. NCLB is currently based on the status approach.

This design issue was central to the paper that my colleague, Charles Clotfelter, and I wrote for my 1996 book (Clotfelter and Ladd, 1996, ch. 2). As we documented at that time using data for 5th graders in North Carolina, status models are not well designed to promote an equity agenda because they inevitably favor the schools with the most advantaged students. This pattern

emerges because of the high positive correlation across schools between the socio-economic status of the students and their achievement. Further, we showed that shifting to a measure based on the annual gains in achievement of individual students reduced the bias toward the advantaged schools. At the same time, though, it did not eliminate it. This is an important point, and one that South Carolina learned back in the mid 1980s. When that state was first designing its growth model, state officials were startled and distressed to find that the growth model still favored the more advantaged schools. To make their proposed accountability system appear fair and, hence politically acceptable, they chose to follow the standard approach used in college athletics, namely to divide schools into divisions – in this case by the socio-economic status of their students – and to reward schools in the top quartile of each of those divisions.

In subsequent work with Randall Walsh, I have explored growth models in more detail paying particular attention to the nature of the incentives they create for more efficient provision of education (Ladd and Walsh, 2002). For a number of reasons including transparency and data limitations, states are not likely to implement a pure value-added model and instead are more likely to focus on gains in student performance without fully adjusting those gains for differences in the socioeconomic status of the students or for the resources available to the school. Though far better than the status measures in terms of the incentives they create for schools to improve student performance, the resulting measures can still have some unintended outcomes. As my Duke colleagues and I have documented for North Carolina, even a relatively sophisticated growth based accountability system can exacerbate the problems that low performing schools face in retaining teachers (Clotfelter et al, 2004). Those adverse effects, however, would undoubtedly be far larger with the especially in small schools

Thomas Kane and Douglas Staiger have correctly raised an additional set of concerns related to measurement error (Kain & Staiger, 2002). Because growth models are based on two test results, not just one, they are subject to more error than status models, and that the errors are likely to be largest in the smallest schools.

So where are we on the status vs. growth approach for judging the effectiveness of individual schools? The status model is appealing because it sends a clear signal that the goal is high achievement for all students. The problem, though, is that simply sending a signal does not assure that the outcome will be achieved, and may well lead to unintended and undesired side

effects If the important values are providing realistic incentives for school improvement, especially for schools at the low end of the performance distribution, the growth approach, though itself somewhat flawed, is clearly preferred to the status model.

3. Accountability has had at best small positive effects on student achievement.

A third area of research focuses on the extent to which administrative accountability systems have succeeded in raising student achievement. One approach is to examine trends on the National Assessment of Education Progress (NAEP), typically referred to as the Nation's Report Card. Figure 1 (below) shows eighth grade math and reading scores from 1990 to 2007 with the vertical line in 2002 denoting the beginning of NCLB. Though 8th grade math scores have been rising since 2002, it is hard to attribute that growth to NCLB since it essentially continues the pre NCLB trend. For reading, the decline in 8th grade scores since 2002 scores is obviously a bit discouraging, except possibly for the recent up tick. Fourth grade test scores, as shown in figure 2, present a somewhat more positive picture but still provide no clear evidence of the success of NCLB.

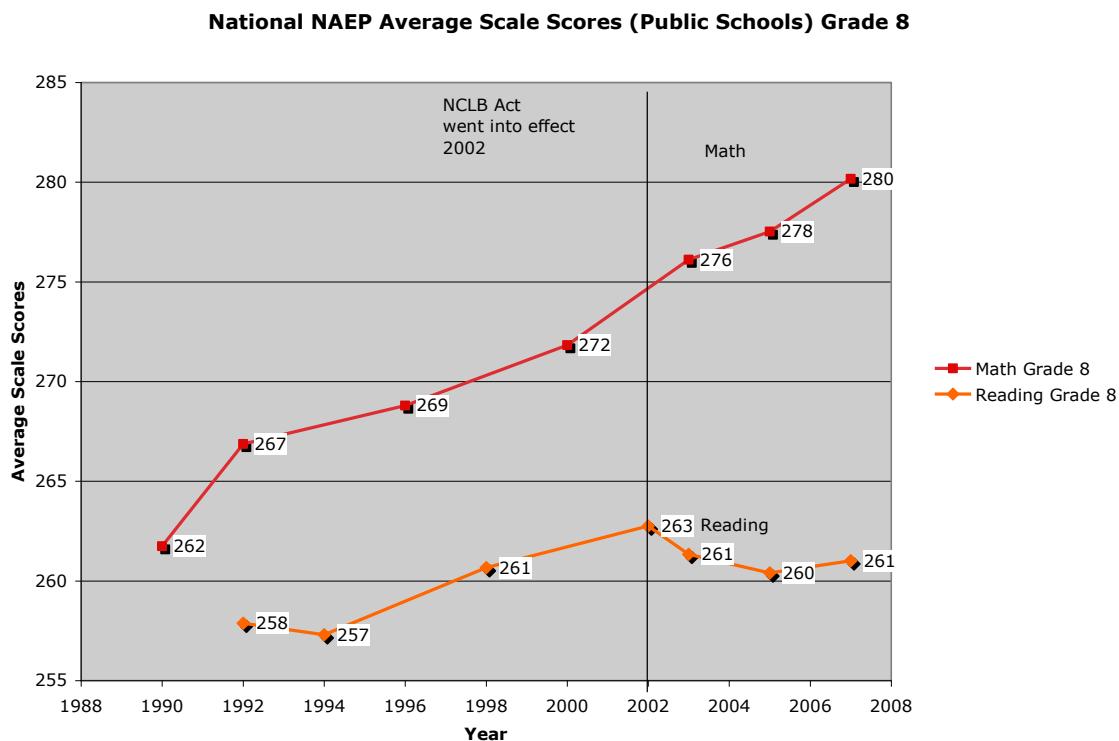


Figure 1

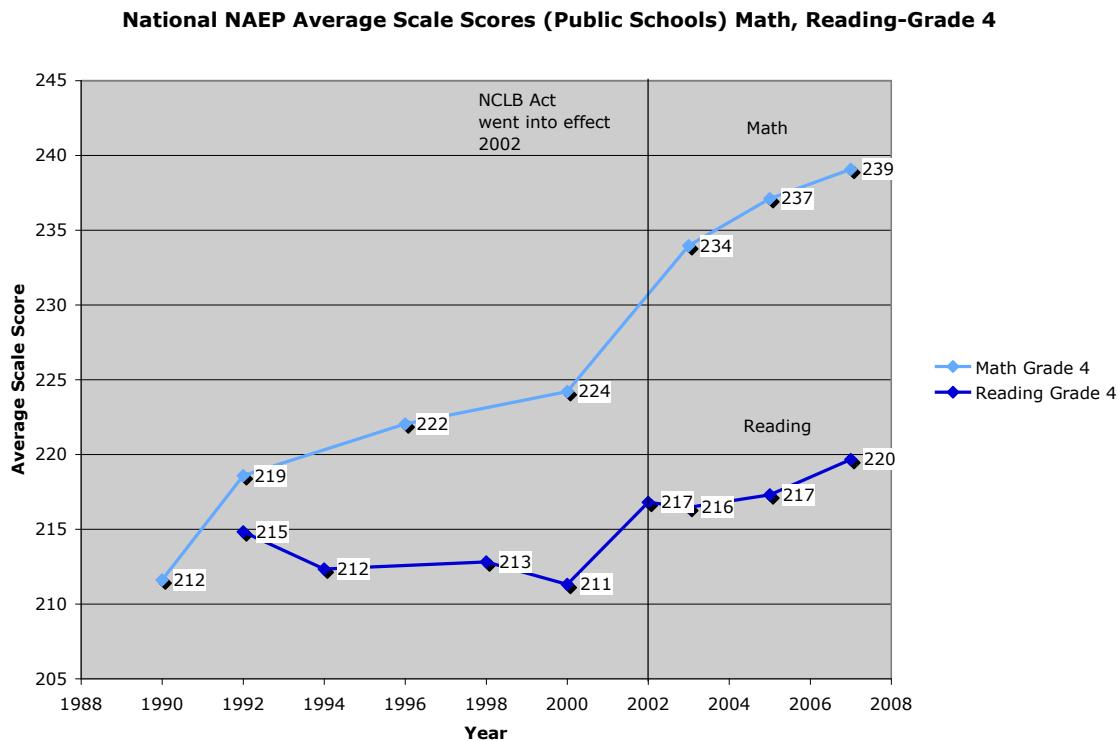


Figure 2

In general, though, it is difficult to determine the effects of a national program such as NCLB. More promising are research strategies that focus on district or state-specific accountability programs or that make use of the variation in accountability systems or in the timing of their introduction across states. In our recent review of the various studies (Figlio and Ladd, 2008), David Figlio and I concluded that Brian Jacob's investigation of the Chicago accountability system (Jacob, 2005), and the cross state studies by Carnoy and Loeb (2002), and by Hanushek and Raymond (2005) are the most methodologically sound.

Three general conclusions emerge from those and the larger set of studies of which they are a part.

First, positive achievement effects of accountability systems emerge far more clearly and frequently for math than for reading. This pattern is particularly clear when the outcome measure is based on a national test, such as NAEP but it also emerges in some of the district or state level studies. One exception to this finding is Jacob's study of Chicago where the positive effects for low performing students were somewhat stronger in reading than in math, at least on the high stakes test, an outcome that may reflect the specific focus on reading in that district . In general

the larger effects on math scores than on reading scores are consistent with findings from other policy interventions such as voucher programs or charter schools.

Second, any positive achievement effects that emerge are quite small. Hanushek and Raymond (2005) found that the introduction of accountability systems with consequences for schools during the 1990s raised eighth grade test scores on the NAEP on average by about 3.2 scale points which is about 1/5 of the cross state standard deviation in scale scores but a much smaller proportion of the cross student variation.

Third, the studies generate mixed results by racial group. Carnoy & Loeb (2002) find larger effect sizes of accountability on passing rates at the basic level on NAEP for black and Hispanic students than for white students. But other studies with different outcome measures find different patterns. In particular, Hanushek & Raymond (2005) find essentially no effects of accountability on the eighth grade achievement of black students, but positive effects for Hispanic students, patterns that are consistent with my own early findings by racial group for 7th graders in Dallas (Ladd, 1999).

The effects of accountability on racial achievement gaps are similarly mixed. Hanushek & Raymond (2005) find that state accountability systems may have reduced the gap for Hispanics but raised it for blacks. And two recent national studies find little effect of NCLB on racially defined achievement gaps (citation needed).

Finally, in a recent study of achievement gaps in North Carolina, my Duke colleagues and I have uncovered a pattern for blacks that suggests that accountability in NC may have raised achievement for blacks at the bottom of the distribution but lowered it at the top of the distribution (Clotfelter, Ladd, and Vigdor, forthcoming).

The bottom line is pretty clear. Accountability has not generated the significant gains in student achievement that policy makers intended, nor is the country on track to meet the high proficiency standards required under NCLB.

Though that law, with its requirement that test scores be disaggregated by racial and income group, has indeed focused new attention on the failings of the system as a whole to meet the needs of many disadvantaged students, it has apparently had limited success in raising the achievement of those students in significant or widespread ways.

What Is Wrong with the Current Approach and Where Should We Go from Here?

NCLB has been criticized for many things, including for requiring unrealistic gains in student achievement and for relying on a concept of proficiency that means different things in different states. Without intending to minimize these and other concerns specific to NCLB, I want to highlight three larger problems with administrative accountability systems that, in my view, have received too little attention but cry out for significant changes in policy focus.

The three problems are: far too little attention to the social and economic contexts that affect educational outcomes, too little attention to the broader education system within which schools operate, and too much focus on test-based accountability

1. Too little attention to the social factors that affect student achievement.

The biggest failing of the current approach to accountability - especially as manifested in NCLB – is the implicit assumption that the education system alone can fully offset the racial and economic disparities that children bring to the school.

Beginning with the Coleman report in the mid 1960s (Coleman et al, 1966), more than half a century of research, both here and abroad has documented a powerful association between social and economic disadvantage on the one hand and low student achievement on the other. Weakening that association is the fundamental challenge facing U.S. education policymakers. But we are kidding ourselves if we think the education system can significantly reduce achievement gaps alone. Though individual schools have shown some success with some students in overcoming social and economic barriers, there is no evidence to suggest that such problems can be sustained and replicated on a large scale.

What is needed is far more attention to high quality interventions in early childhood, greater investments in health services for pregnant mothers, infants, and for children as they progress through school, and greater attention to out-of-school enrichment opportunities for economically disadvantaged children. Janet Currie has shown for example that about one quarter of the racial achievement gap in school readiness can be attributed to a combination of racial differences in health and maternal health and health behaviors (Currie, 2005, p. 133). And Nobel prize winner James Heckman has described investments in disadvantaged young children

as a “rare public policy initiative that promotes fairness and social justice and that at the same time promotes productivity in the economy and in society at large” (Heckman, 2006, pp. 1902).

Of course redirecting education policy making toward these contextual factors is often hard to do in practice. It is not easy to induce multiple government agencies, each their own separate culture and funding sources, to work together.

I am hopeful, though, that improved research ultimately will be useful in pushing the policy discussion and policy practice in a new and more productive direction. My hope is that in the not too distant future, researchers will figure out a way to link data on early childhood experiences to longitudinal administrative data sets on students in schools so that we will be able to document more clearly for policy makers the effects of various early childhood and health interventions on subsequent educational outcomes. That would provide empirical support for a policy strategy that extends beyond the schools themselves for reducing achievement gaps by race and family income.

Nonetheless, we still need to pay attention to the education system itself, which takes me to the second major failing of the current accountability movement.

2. Too little attention to the broader education system within which individual schools operate.

Although some states, including North Carolina, initially focused their test-based accountability systems on school districts, the main focus of accountability systems is now the individual school. This focus is part of a larger trend in the U.S.– and also in countries around the world – to shift managerial responsibility to the school level (Plank & Smith, 2008). The standard arguments for focusing attention on individual schools rather than districts are that the school is where the rubber hits the road, and that the focus on schools minimizes the chances that poorly performing schools within an otherwise high performing district will escape public scrutiny. Now under NCLB, states have no choice. The school is the object of attention.

But it is time, in my view to rethink the punitive pressure currently being placed on individual schools. For many purposes, the school is not the right unit. Consider the implications under NCLB of holding each school accountable for the performance of subgroups defined by race and income within the school. Because of small sample sizes within schools, states have set floors on the number of students a school must have in a subgroup for the subgroup to be

reported separately. Those floors then provide strong incentives for district policy makers to distribute students among schools to keep as many of the historically low-performing students below the floor as possible. The result is that millions of students are not included in measures of subgroup performance. An alternative would be to hold districts, not schools, responsible for the performance of subgroups. Or think about what happens when students move in and out of schools during the year. Clearly it is not fair to hold a school accountable for a student who is in the school for only a short time, but who then is accountable for such students?

More generally, because schools are part of a larger education system, they do not have control over some key aspects of their environment such as the resources available to them. And it is district policy makers – not individual schools – who set the rules and design the incentives that determine how both students and teachers are distributed among schools. Further, with the rising enthusiasm for small schools, schools within schools, charter schools, and contract schools, it is not always fully clear what we mean by a school. Within each district we have a system of schools. Yet under the current approach to accountability, we place almost all the blame for student failure on educators in individual schools.

My own recent work on teachers in North Carolina has increased my sensitivity to this issue. Over time, in that state, children from low income families have become increasingly concentrated in schools with large proportions of poor children. (Clotfelter, et al, 2007). That is cause for policy concern for many reasons but especially because our evidence also clearly shows that schools with large proportions of poor children end up with lower quality teachers than richer schools. And with the federal courts now out of the business of playing an active role in promoting racial desegregation we are seeing rising rates of racial segregation throughout the state, but perhaps most distressingly so in Charlotte-Mecklenburg which once was the poster child for racial desegregation. With the federal courts no longer providing political cover for local leaders who would like to keep schools from resegregating, the situation will only worsen. Yet segregation matters for student achievement because it is typically low income black or Hispanic children who end up in the racially isolated schools, and it is those schools that our research documents, as does that of others, that face the greatest challenges in attracting and retaining high quality teachers (see, for example, Boyd et al, 2008).

So, at a minimum, we need to stop placing all the blame on the teachers and principals in specific schools. Where is the accountability for other policy makers, including state legislators or county commissioners who provide insufficient resources, local school boards who do little to counter the strong pressures for income or racial imbalance of students across schools within districts, or state and district policy makers who determine the rules and incentives under which teachers are distributed among schools and who often fail to provide the support services needed to make the schools work better?

The pressure on those groups will ultimately have to come through the political process, perhaps in some cases with a nudge from the courts, and will require strong leadership at the state and district levels. Researchers, I believe can play a positive role by providing the evidence needed by policy makers to make the case for this more systemic focus and by helping to evaluate policies designed to even out the distribution of quality teachers across schools.

The question still remains, however, about the appropriate form of accountability for individual schools. That leads me to my final failing of current accountability efforts.

3. Too much reliance on test-based accountability— and too little attention to promoting effective process and practice within schools.

As someone who is continually struggling to keep my weight in check, I fully understand the need for having a scale in the bathroom, and believe in the maxim that if you don’t measure it you won’t do anything about it. Analogously, I believe testing students on a regular basis is an essential component of a well functioning education system. But too much reliance on high stakes tests for accountability can be counterproductive when it is used primarily for punitive purposes.

I have already referred to the narrowing of the curriculum that inevitably occurs when schools are held accountable for student performance in only a few subjects. Of course, one solution to that problem might be to test more subjects, including for example, science, social studies, art, and languages, but we then run into two very real constraints. One is the time constraint – the more time schools devote to state wide testing the less time there is for instruction. The other is the demands on a testing industry that is already overwhelmed with the existing demands for curriculum referenced tests (Toch, 2007).

A second concern about tests, and one that is generating much discussion within the North Carolina Blue Ribbon Commission on Testing and Accountability, is the fact that tests are not very effective at evaluating, and hence promoting, 21st century skills such as problem-solving, team work, and collaboration within diverse environments. Ultimate employers are not looking for students who can fill in bubble sheets on tests. Instead they are looking for students who can function effectively in a global and fast-changing world.

A third set of concerns is the stress that testing can impose on students as they face increasing pressure to achieve, and the loss of morale among teachers as they struggle in many cases, particularly under the requirements of NCLB, to meet unrealistic objectives. To the extent that both the student stress and the low teacher morale are connected with unrealistic expectations, or, in the case of teachers, with insufficient capacity or knowledge about how to proceed, more test-based accountability is not going to improve the situation.

A fourth concern, and one that I want to highlight, is that the focus on test scores by itself does little to promote good practice. Though some limited evidence is now emerging from Florida that its test based ratings of schools has led to some improved practices, my sense is that accountability can and should be used in a more proactive way to promote good practice, but with full recognition that good practice varies across school contexts..

Given my values, here is my vision of a more balanced accountability system. States would still use test results to hold schools accountable for realistically obtainable gains in student performance in core subjects such as math or reading at the elementary level, and for some basic subjects in high school. That type of accountability would be supplemented by a new system of school inspections designed to improve practice and to encourage schools to pursue many of the other outcomes demanded of a good education system.

This vision is consistent with my basic view that a good accountability system is one that is constructive and not punitive. The specific idea was inspired by my research on the Education Review Office in New Zealand (Fiske and Ladd, 2000, ch. 5). That office was set up in the early 1990s to monitor the performance of that country’s schools after the government decentralized operating authority to the school level. Related approaches can be found in the audit or reaccreditation procedures at the university level in the U.S. and other English speaking countries, and, in a related field, in the form of national oversight of hospital quality.

Although the details would need to be worked out, here is generally what I have in mind. Each state would have a statewide review board that would be independent of the State Board of Education. The review board would then send small teams of professionals to make periodic visits to each school – perhaps one visit every two or three years – with each visit preceded by an internal self-study. The review panel would then write a report on each school that, along with the school’s response, would be made public. Though the report would include a summary of the school’s success, or lack thereof, in raising student achievement in the core subjects, the report would evaluate the school on a far broader set of outcomes than student test scores alone.

The ultimate concern would still be student outcomes, but individual schools could define for themselves which additional outcomes were most important. Moreover, the review panel would look closely at the policies and systems that schools put in place to promote those outcomes. The review panel itself would not be in the business of providing assistance or support to the school since doing so would interfere with its ability to be objective.

This approach is designed to do the following. First and foremost, it would force schools to reflect on what outcomes, besides simply those that are targeted under the state’s high stakes accountability system, they are seeking to promote; what strategies they are using to achieve them; and what data they are using to measure their success at reaching those goals. Ironically, the result could well be more testing of students, not less, but with the tests being used more for internal diagnostic purposes within the classroom than for school based accountability. The intent here is to encourage the schools to develop their internal capacity to make data-driven decisions, while not forcing them into a straight jacket of common outcomes and practices.

At the elementary level, goals might include higher achievement in a broader set of subjects than the ones tested by the state, as well as intermediate or process goals such as greater participation of students in the arts. Included among the goals at the high school level might be intermediate outcomes such as a higher percentage of minorities enrolling in advance placement or other advanced courses, or the provision of opportunities for more students to work effectively in teams.

In addition, the approach would identify the challenges the schools face in meeting those goals. Though the review panel itself would not provide assistance to the school, the underlying concept is that either the district or state policy makers would make use of these reports in

allocating resources, providing technical assistance, or otherwise making sure schools have access to the resources. In some cases higher level policy makers might use the information from the reports to work with schools to raise their aspirations.

Because the review board would ultimately be looking at all schools, over time it would uncover a number of different strategies used by the schools to achieve particular goals. The review panel might then write periodic overview reports describing the types of strategies used by different schools and drawing on nationwide scientific research to document the potential for the various strategies to be successful. In a sense this activity fulfills the same role as the formal benchmarking processes used by business firms such as Xerox. In the business context, individual firms devote significant amounts of their own resources to learn about the best practices of other firms. Because schools do not have the luxury of making such investments, the review board could fill that gap by learning about and publicizing information about best practices related to particular goals within similar types of schools, information that would then be available as a public good to all schools.

One might well ask where the personnel for such review panels would come from. and how much such a system would cost. Those are obvious key questions and the most I can do at this point is to suggest some answers. With respect to the first, strong leadership for the review board will be essential and panel members must be knowledgeable, have experience with education, and have good judgment. One obvious concern is that the review panels might pull the best teacher and administrators out of the schools, to the detriment of student learning within the schools. If, however, some of the positions were relatively short term – say three years—one could imagine using the review board it as a way to provide additional short term career opportunities for highly qualified principals and teachers after which they would return to the schools. In this sense, service on the board could be viewed as a professional development opportunity that would benefit not only the members, presumably in the form of higher pay for that period, but also the schools to which they returned.

To be sure, such an inspection system could potentially be costly both to the state and to the schools being reviewed. The approach is appealing to me, however, because I view it as a far more positive and constructive form of school accountability than the current system. The relevant policy question is whether the benefits, many of which are in the form of promoting

better decision making processes within the schools, expanding information about best practices, and sending signals that schools have a broader agenda than simply improving math and reading, are worth the costs. My own view is that there is a good chance they are and hence it is time to start serious discussion of models of this type.

Conclusion

To conclude, I return to a central theme of this lecture, namely the complex interrelationships between values, policy and research. When I first entered the field of education policy research in the mid 1990s after more than 20 years doing research on state and local public finance, I was amazed – and distressed – at the ideological content of some of the research. Some education policy researchers appeared to know the best policies even before they did the research.

One of the conclusions that emerges from this discussion of school accountability is that values are deeply embedded in education policy. Hence, even if they want to, policy researchers cannot ignore the role of values. So what is the role of values in good policy research?

My view is that we as researchers will be most effective in promoting good policy if we choose research programs and topics that are consistent with our values, but then maintain the highest possible standards of objectivity in doing that research, and clarify for policy makers the extent to which our policy advice emerges from our values or from our research findings.

In that spirit, I hope that those of you who share my values and conclusions about the failings of the current approach to school accountability will pursue some of the research themes I have suggested. That would include exploration of the effects of policy interventions in early childhood and children’s health on educational outcomes, evaluation of policies to reduce the inequitable distribution of students and teachers across schools, and serious exploration of the benefits and costs of a more balanced accountability system for individual schools.

References

- Bryk, A. S. and K.L. Hermanson. 1992. “Educational Indicator Systems: Observations on Their Structure, Interpretation, and Use.” *Review of Research in Education*, 19: 451-84,
- Boyd, D. , H. Lankford and J. Wyckoff. 2008 “Increasing the Effectiveness of Teachers in Low-Performing Schools.” In H.F. Ladd and E.B. Fiske, *Handbook of Research in Education Finance and Policy*. Lawrence Erlbaum/Routledge Press.
- Carnoy, M. and S. Loeb. 2002. “Does External Accountability Affect Student Outcomes? A Cross-State analysis.” *Educational Evaluation and Policy Analysis* 24(4), pp. 305-331.
- Clotfelter, C.T. , H.F. Ladd, and J. Vigdor, Forthcoming. “The Academic Achievement Gap in Grades 3-8. *Review of Economics and Statistics*.
- Clotfelter, C.T. , H.F. Ladd, J. Vigdor and R. Aliaga-Diaz, 2004. “Do Accountability Systems Make It More Difficult for Low-Performing Schools to Attract and Retain High Quality Teachers?” *Journal of Policy Analysis and Management*, 23, no. 2. pp. 252-271.
- Clotfelter, C. T. and H.F. Ladd. 1996. “Recognizing and Rewarding Success in Public Schools. “ In H. F. Ladd, editor, *Holding Schools Accountable: Performance Based Reform in Education*. Washington, D.C.: Brookings Institution.
- Clotfelter, C.T. , H.F. Ladd, J. Vigdor, and J. Wheeler. 2007. “High Poverty Schools and the Distribution of Teachers and Principals,” *University of North Carolina Law Review*. Vol. 85, no. 5, pp. 1345-1380.
- Coleman, J.S, E.Q. Campbell, C.J. Hobson, J. McPartland, A.M. Mead, F. D. Weinfield, and R.I. York. 1966. *Equality of Educational Opportunity*. Washington, D.C. U.S. Department of Health, Education, and Welfare.
- Cullen, J. and R. Reback. 2006. “Tinkering Towards Accolades.: School Gaming Under A Performance Accountability System.” In T. Gronberg and D. Jansen, eds., *Improving School Accountability: Check-Ups or Choice, Advances in Microeconomics*, vol 14, pages, 1-34. .
- Deere, D. and W. Strayer. 2001. “Putting schools to the test: School accountability, incentives and behavior.” Working paper, Texas A&M University.
- Currie, J. “Health disparities and gaps in school readiness.”. *Future of Children*. Vol 15, no. 1, pp. 117 -138.
- Figlio, D. (2006b) “Testing, Crime and Punishment.” *Journal of Public Economics*.

Helen F. Ladd, “Holding Schools Accountable Revisited”
Page 20 of 21

Figlio, D. and L. Getzler. 2006. “Accountability, Ability and Disability: Gaming the System?” In T. Gronberg and D. Jansen, eds., *Improving School Accountability: Check-Ups or Choice, Advances in Microeconomics*, vol. 14, pp. 35-50. .

Figlio, D. and J. Winicki. 2005. “Food for Thought? The Effects of School Accountability Plans on School Nutrition.” *Journal of Public Economics*.

Figlio, D. and H. F. Ladd. 2008. “School Accountability and Student Achievement.” In H.F. Ladd and E. B. Fiske, *Handbook of Research in Education Finance and Policy*,. Lawrence Erlbaum/ Routledge Press.

Fiske, E. B. and H.F. Ladd. 2000. *When Schools Compete: A Cautionary Tale*. Washington, D.C.: Brookings Institution.

Gibbons, R. (1998). “Incentives in Organizations,” *Journal of Economic Perspectives*, 12(4), 115-132.

Hanushek, E. 2006. “Policy Analysis: Is It, or Could It Be, the Fifth Estate?”
www.appam.org/awards/pdf/2005Spencer_Hanushek.pdf.

Hanushek, E and M. Raymond. 2005. “Does School Accountability Lead to Improved School Performance? “ *Journal of Policy Analysis and Management* , vol. 24, no. 2, pp. 297-329

Heckman, J. J. 2006. “Skill Formation and the Economics of Investing in Disadvantaged Children.” *Science*, vol. 312 (June), pp. 1900-1902.

Jacob, B. 2005. “Accountability, Incentives and Behavior.” *Journal of Public Economics*.

Kane, T. and D. Staiger. 2002. “Volatility in School Test Scores: Implications for Test-Based Accountability Systems.” In *Brookings Papers on Education Policy, 2002*, D. Ravitch, ed. Washington, D.C.: Brookings Institution, pp. 235-83. .

Ladd, H.F. 1996. *Holding Schools Accountable: Performance Based Reform in Education*. Washington, D.C.: Brookings Institution.

Ladd, H. 1999. “The Dallas School Accountability and Incentive Program: An Evaluation of its Impacts on Student Outcomes.” *Economics of Education Review*.

Ladd, H. and Walsh, R. 2002. “Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right.” *Economics of Education Review*, 21(1), pp.1-17.

Ladd, H. and F. Zelli. 2002. “School-Based Accountability in North Carolina: The Responses of School Principals” *Education Administration Quarterly*.

Milgrom,P. and J. Roberts. 1992 *Economics, Organization, and Management*. Prentice Hall.

Smith M.S. and J. O'Day 1991. "Systemic School Reform." In S. H. Fuhrman and B. Malen, eds. *The Politics of Curriculum and Testing: The 1990 Politics of Education Association Yearbook*. New York: Falmer Press. Pp. 233-267.

Plank, D.N. and BA Smith. 2008. "Autonomous Schools: Theory, Evidence and Policy" In H.F. Ladd and E. B. Fiske, *Handbook of Research in Education Finance and Policy*,. Lawrence Erlbaum/ Routledge Press.

Toch, T. 2007. "Margins of Error: The Education Testing Industry in the No Child Left Behind Era." www.educationsector.org/research_sho.htm?_id=346734. Accessed 9/11/07.