Revised
March 30, 2004

Policy Brief on Accountability

for

Symposium on School Finance and Organizational Structure New York State

Helen F. Ladd
Edgar Thompson Professor of Public
Policy Studies and Professor of Economics
Sanford Institute of Public Policy
Duke University
Durham, NC 27708

919-613-7352
Helen.ladd@duke.edu

Abstract

This policy brief examines some of the design and incentive issues associated with North Carolina's widely praised school-based accountability program, with additional attention to its relationship to the federal No Child Left Behind Act of 2001. The North Carolina accountability system is embedded in a standards-based approach that starts with a statewide standard course of study, annual state tests that are aligned with the state curriculum, recognition and financial rewards for successful schools and interventions for unsuccessful schools. One set of issues relates to the measurement of school effectiveness. By rewarding teachers for gains in student performance that are based on annual test scores for all students in grades 3-8, North Carolina minimizes, but does not eliminate, the possibility of simply rewarding teachers who happen to be teaching in schools serving more affluent students.

A key lesson for policy makers is that accountability systems are powerful tools for changing the behavior of school personnel in both intended and unintended ways and hence must be designed carefully. Surveys of North Carolina school principals indicate the power of the accountability system to change their behavior in ways that are generally consistent with state goals. In addition, however, the survey responses illustrate the potential for resources and attention to be shifted away from subjects not targeted by the accountability system in favor of math and reading. Furthermore, the survey responses as well as additional analysis of the movement of teachers indicate that the accountability program has made it more difficult for low-performing schools to retain teachers.

The subgroup requirements of the Federal No Child Left Behind NCLB) Act pose new challenges for state policy makers. Though such requirements apply at the school level, this brief makes the case for their use at the district rather than the school level. Finally, state policy makers should be under no illusion that the choice provisions of NCLB will generate greater achievement in the schools to which they apply.

Accountability in education is a broad concept and could be addressed in many different ways: through the design of political systems to assure democratic accountability to voters, the introduction of market-based reforms to increase accountability to parents and children, the development of peer-based accountability issues to increase the professional accountability of teachers, or the use of top-down administrative accountability systems designed to drive the system toward higher student achievement.

This policy brief focuses attention of this last form of accountability. Specifically, it uses North Carolina's experience with an administrative school-based accountability system to draw some general policy lessons about such programs. North Carolina's experience is of interest in part because its accountability system is relatively sophisticated and has been in place for several years and in part because it served, along with Texas, as a model for the accountability requirements of the Federal No Child Left Behind (NCLB) Act of 2001.

Implicit throughout this discussion are two underlying assumptions. One is that adequate funding for schools is a prerequisite for the fairness and success of an accountability system. Without adequate funding, schools may not have the resources needed to achieve the goals of the program. The other is that such systems work best when they are designed to be constructive and positive rather than punitive. Even if the fear of being sanctioned served as a motivating force for higher student achievement in low-performing schools in the short run, it would be hard to sustain such improvement over time given that good teachers and school principals have options other than to teach in such schools. Instead, school accountability systems are better used to provide positive incentives for greater student achievement and to identify problems in particular schools that those problems can be addressed in a positive and constructive manner.

This brief does not cover all issues related to accountability systems. Instead it focuses on a select set of issues about which I have direct evidence from North Carolina.The first section describes the North Carolina ABCs program with attention to the characteristics that make it distinctive. The second section argues that school-based accountability is a powerful tool for changing behavior, in both intended and unintended ways. The third and fourth sections look at two specific aspects of the No Child Left Behind Act: the requirement that test scores be disaggregated by subgroup and that parents of children enrolled in schools that are failing be offered the choice of another public school.

1. North Carolina ABCs Program

In contrast to many states in the Midwest and the Northeast , including New York, the state government in North Carolina has played a large role in education policy and school finance ever since the depression when the state essentially took over responsibility for education from poverty stricken school districts. In particular the state

takes responsibility for financing a basic level of education for students in all of the state's districts, most of which are conterminous with the state's 100 counties. Consistent with this large state role, the state imposes a standard course of study and a statewide salary schedule for teachers. Local districts can supplement education spending and teacher salaries to the extent that county governments are willing to provide local tax revenue for that purpose Currently, the state of North Carolina provides about 70 percent of the state and local share of revenue for schools.

The following paragraphs highlight the process that led to the state's accountability system, the sophistication of its design, and the fact that it appears to be having a positive effect on student performance.

Evolution of the ABCs program over time and as part of a broader reform effort. The state's highly touted accountability system, the ABCs program -- where A stands for accountability, B for basic skills and C for local control--- is part of a broader reform effort in North Carolina to improve the academic performance of the state's children. During the 1990s under the leadership of Governor Hunt the state raised teacher salaries from far below the national average to the national average, invested in a variety of programs to improve the training of teachers and has promoted children's readiness to learn through the governor's award-winning Smart Start initiative.

As part of that reform effort, the ABCs program was carefully developed over an extended period of time with the support of the business community and the buy-in of policy makers and many school officials. In the early 1990s, the state introduced an innovative district-level reporting system designed to hold districts accountable not for their actual performance but rather for their performance relative to the performance that would be expected given the profile of the students they served. Consistent with the national standards movement with its focus on student performance relative to a standard, in 1992-93, the state introduced curriculum-referenced tests that are aligned to the state's standard course of study. Proficiency, or grade-level performance, in each grade was deemed to be III or IV on a four point scale.

Since then the state has been testing all students annually in grades 3-8 in math and reading and in writing in grades 4 and 7, and students at the high school level at the end of courses such as algebra I, English I and II, U.S. History and Biology. The ABCs school- based accountability program was introduced in 1996-97 for elementary and middle schools based on the end-of-grade test scores for students in grades 3-8 and the following year for high schools based on the end-of-course tests by subject. Under the ABCs program, the teachers and staff in schools deemed to be effective in raising student achievement are given financial bonuses and those that are deemed to be low performing are subject to state intervention. In the years between 1996 and the present, the State Board of Education has made minor modifications to the program almost every year in an effort to improve it and in response to changing circumstances. The biggest recent changes are the efforts of the State Board to incorporate into the ABCs program the requirements of the federal NCLB legislation.

Good design with focus primarily on gains in test scores. A major design issue with respect to any school based accountability program is whether to focus on the levels of student performance or on gains in that performance. For the purposes of financial incentives, North Carolina emphasizes gains rather than levels. Given that a large proportion of the schools are now meeting their expected gains, the state currently appears to be placing somewhat more emphasis on levels for purposes of public recognition. (See Table 1 below for the designations based largely on levels of performance.)

Central to the ABCs program and a crucial contributor to its initial acceptability is the fact that the financial rewards for school personnel are based exclusively on the average gains in test scores for the cohort of students in the school during the year. For each school serving students in grades 3-8, an expected average gain in test scores is predicted and the school is deemed effective or not depending on how the actual gains of its students compare to its predicted gains.[1] This approach is feasible only because the state annually tests all students so that the gain for each individual student can be calculated. The goal of the program is to provide incentives for each school to provide its students with a year's worth of learning for a year's worth of education.

This cohort gain approach is far superior to two alternative approaches for measuring the effectiveness of schools: use of levels (such as the average test score or the percent of students who are proficient) or changes in average outcomes (such as test scores or the percent proficient) from one year to the next. The use of levels is undesirable because test results are highly correlated with the socio-economic status of the students in the school. Thus when levels are used the high positive correlation between a school's effectiveness ranking and the affluence or socioeconomic status of the students it serves means that teachers in the more affluent schools are privileged relative to other teachers regardless of their contribution to student learning. The change in average test scores from one year to the next is flawed because that change could well reflect year-to-year changes in the profile of the school's students rather than gains in performance by individual students. The gain measure avoids that problem because it is based on the gains of a specific group of students. It is, however, subject to the criticism of being imprecise because of the chance variation in test results, the effects of which are exacerbated by the need to look at two years of data. .

---

[1] The expected average gains are predicted as follows. For each grade and subject (i.e. math and reading), a student's expected score is based on an equation of the form $TS_t - TS_{t-1} = a + bX1 + cX2$ where TS is the test score in either math or reading in year t and TS the test score in the same subject in year t-1, X1 is a proxy for the student's proficiency and is measured as the sum of the student's math and reading scores for the previous year minus the state average, and X2 is designed to account for regression to the mean and is measured as the student's prior year score in the subject of interest minus the state average in that subject. The tests are scored on a developmental scale and the parameter "a" can be interpreted as the statewide average gain in score for students in the specified grade and for the specified subject. The parameters a, b, and c were estimated using 1994 test scores for each grade. Because the b and c coefficients were quite similar across grades for each subject area, the state uses a single pair of b and c coefficients for each subject area to determine the expected growth rates. For further discussion see Ladd and Walsh, 2002.

Though in many ways it is the best of the three approaches, the gain measure used by North Carolina still does not measure the true value–added by the school, where value added is defined to be the contribution of the school's staff, as distinct from the characteristics of the student body, to the learning of its students. To calculate a true value added measure, policy makers would have to adjust any average gains by the mix of the students in the school, something that is not generally done, and also for the resources available to the school. As a result, the effectiveness of schools as determined by the average cohort gain approach is still positively related to the socio-economic characteristics of the school, albeit far less so than would be the case for average test scores or the percent of students at a proficient level. As a consequence, teachers in schools serving more advantaged students still have a somewhat higher probability of winning a bonus than do teachers in schools serving less advantaged students.

This distinction between gains in scores and true value added measures of school effectiveness has a direct policy implication. It means that the gain measure should be used more for the purpose of driving change than as a vehicle for placing blame. To the extent that a school appears ineffective based on its gains in student performance, the cause could be the lack of insufficient resources for the school given the particular mix of students it serves. Stated differently, any shortfall in student performance should not necessarily be attributed to the teachers alone.

For the first few years of the ABCs program, schools were given one of the following four rankings: 1)exemplary, 2) meets expectations, 3) no recognition, and 4)low performing. In 2001 the term exemplary was replaced by high growth. Exemplary or high growth schools are those whose gains in student test scores exceed the predicted gains by more than 10 percent. Teachers in those schools currently receive bonuses of $1500 and teaching assistants receive bonuses up to $750. The schools that meet expected growth are those whose gains meet the expected gains but exceed them by less than 10 percent. Teachers in those schools receive bonuses of $750 and assistants receive $375. Some of the schools that fall into the high growth or expected growth categories, and hence whose teachers are eligible for awards, are now labeled priority schools on the ground that less than 60 percent of their students are performing at grade level.

At the bottom are schools receiving no recognition and low performing schools. Schools in neither of these groups meet their expected growth targets. The difference between no-recognition schools and low-performing schools is that in the latter, more than half of the students are below grade level.[2] Thus, the low-performing schools meet neither their school specific growth standard nor the state wide performance standard of at least 50 percent at or above grade level. As noted above, a school is deemed to be a priority school is less than 60 percent of its students are at grade level, regardless of its growth performance.

[2] Any no-recognition school with less than 60 percent of its students at grade level is labeled a priority school, along with all other schools (other than low-performing schools) with less than 60 percent of the students at grade level. .

State assistance teams are designated to work with the low-performing schools to improve their performance. State law requires the team to conduct a comprehensive needs assessment, to evaluate teachers and administrators, and to help the school revise (if necessary) and implement its school improvement plan. In addition, the state makes available voluntary assistance teams to no-recognition schools and priority schools. Fifteen schools received mandatory state assistance in 1997-98 but only 5 schools did so in 2003.

Rising student achievement. Recent gains in state administered tests suggest to some observers that the ABCs program is having strong positive effects on student achievement. As reported by the North Carolina Department of Public Instruction, the percent of all students in grades 3-8 scoring at or above grade level rose from 53 percent in 1993 to 75 percent in 2002 Consistent with this observation is the declining share of low-performing schools as shown in figure 1 and the large increase in schools meeting the high growth standard. While somewhat less than 60 percent of the schools were meeting their expected or high growth standards in 1997, 94 percent were doing so in 2003. During that same period the number of low-performing schools decreased from 6.7 percent to less than half a percent.

Further evidence of the apparent success of the program emerges from North Carolina's rising scores on the National Assessment of Educational Progress (NAEP) scores.[3] While those gains were generally larger than those in the nation as a whole and also than those in New York, they were not consistently higher than those in Florida, another state with a long established school-based accountability system. Shown in figure 2 are increases in NAEP scores relative to the national average for the three states. While all three states exhibited gains in NAEP scores that were comparable to or greater than those for the country as a whole, North Carolina students exhibited gains of comparable magnitude to those in Florida in 4th grade math, and the largest gains of the three states in 8th grade math. Florida exhibited the greatest gains in reading.

Although it is tempting to attribute the above-average test-score gains in NC to the state's accountability program, one must be careful in doing so since accountability was only one part of a larger reform effort in the state. Instead, the gains should be attributed to the overall reform effort. More general support for the conclusion that strong accountability programs promote higher achievement emerges from the recent research of Carnoy and Loeb (2002) who perform a cross state analysis of the power of accountability systems to raise student performance on the NAEP. They find that the students in strong accountability states exhibited greater gains in test scores between 1996 and 2000 than did students in states with weaker accountability systems. In addition, they find no evidence that strong accountability raised retention rates or reduced high school completion rates. [4]

[3] The changes in math scores cover the period 1996-2003. The changes in reading scores cover the period 1998-2003.

[4] North Carolina and Florida are both included among the high accountability states. In the basic analysis New York is placed in that category as well, but in additional analysis, New York is given a much lower

.

II. The Incentive Effects of School-Based Accountability.

The main purpose of North Carolina's school based accountability program is to alter the incentives facing school personnel in such a way as to increase student performance. The program operates by changing the focus of accountability away from the traditional emphasis on inputs toward student outcomes, giving the schools more flexibility over operations than they previously had, and providing incentives for schools to meet expected outcome goals. In contrast to incentives directed at the performance of individual teachers that may induce competitive rather than cooperative behavior, a major justification for focusing accountability at the school level is that, such programs provide incentives for school personnel to work together in a cooperative manner to achieve the desired goal of higher achievement. In addition, compared to individual teachers for whom the main behavioral response would simply be to exert more effort, schools have a wider range of potential responses provided, of course, that they have been granted managerial flexibility over school operations.

Educators are rightly concerned about a number of incentive effects that are not elaborated in this section, including, for example, the incentive for teachers to teach to a specific test or to narrow the curriculum. A full evaluation of any school based accountability system would have to pay close attention to those effects as well as the incentive effects discussed in the next two sections: the responses of school principals and the movement of teachers among schools.

Effects on the behavior of school principals. The ABCs program undoubtedly provides powerful incentives for one set of key school actors, school prinicipals, to change their behavior. Evidence for this assertion comes from a recent study of the responses of elementary school principals to the ABCs program (Ladd and Zelli, 2001). Our analysis was framed within the context of a principal-agent model in which we view state policy makers as the "principal" with the goal of improving educational outcomes in the state and the school principals as the "agents" of the state. We were interested in the extent to which North Carolina's ABCs school accountability system has succeeded in altering the behavior of school principals in the direction desired by the state. To that end, we surveyed principals both about their attitudes toward the ABCs program and about how they have altered the way they allocate their own time and the use of resources within the school. We administered the survey first in the summer of 1997 after the program's first year and again in the summer of 1999. The survey was based on responses from about 65 school principals randomly chosen from schools across the state.

The school principals generally supported the ABCs program, with over 60 percent having a positive view overall (at the level of four or five on a five-point scale) and the others being evenly divided between a neutral and a negative view of the program. This positive attitude toward the program, and also the principals' general

---

accountability rating based on the fact that its accountability system is so recent. See footnote 14 in Carnoy and Loeb (2002).

agreement with many of the basic components of the program, is largely attributable to strong educational leadership at the state level mentioned earlier and to the state's efforts to communicate with local school officials. Further, it reflects the fact that the ABCs program itself was part of an accountability movement that had been evolving throughout the decade.

The survey results indicate that school principals were well aware that student performance on the end-of-grade tests would play a larger role than in previous years in the direct evaluation of their own performance and also would affect the rating of their schools. In addition, two out of three principals perceived that the ABCs program had increased their ability to make their teachers more effective, presumably in part by highlighting the potential for teachers to earn financial rewards or public recognition. Without such empowerment, principals could well have believed that they did not have the tools necessary to effect change at the classroom level.

The ABCs program clearly changed the behavior of school principals. By comparing the actions the respondents described as ongoing at the time of the 1997 survey to the actions described as ongoing and new in the 1999 survey, we were able to conclude that principals had become much more active in a wide range of policy areas. In response to the ABCs program, we found that school principals increased their use of end-of-grade tests as a diagnostic tool to help teachers improve instruction, developed extracurricula programs focused on math and/or reading, spent more time with teachers on classroom instruction, and encouraged greater focus on math and reading in the teaching of other subject areas. All of these actions are fully consistent with the state's goal of raising student achievement. In addition, principals redirected resources from other subject areas to math and/or reading and encouraged teachers to spend more time on the teaching of test-taking skills, new actions that the state did not necessarily intend and that in some ways could be detrimental to students.

Importantly, the incentives of the program were sufficiently strong to induce even those school principals who opposed the ABCs program to change their behavior in ways similar to the supporters of the program. While principals who agreed with the basic thrust of the ABCs program were initially more aggressive than the others in changing their behavior, by 1999, the actions of principals who did not support the basic thrust of the ABCs program were indistinguishable from those who did.

Although many of the reported actions may be desirable, 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 other goals (Gibbons, 1998 and Milgrom and Roberts, 1992, pp. 228-31.) Many educators are implicitly evoking this theorem when they express fear that high stakes accountability systems threaten the complex fabric of education.

The difficulty of creating ideal measures of performance when there is more than one task to be performed is a standard problem in the principal-agent literature. Gibbons (1998) proposes two solutions to the problem. One is to keep any incentives that are based on imperfect measures of overall performance relatively small. In the context of educational accountability that might be achieved by limiting financial bonuses for school success on state tests, which represent only one of several outcomes the state ultimately values, to a small percentage of the salary levels of teachers and principals. The other is to use multiple instruments to provide a balanced package of incentives. In the education context, that might be achieved by evaluating school principals not only on the basis of their school's short run performance on state tests but also on the basis of other outcomes longer run such as their success in developing effective teachers and maintaining a nurturing school environment.

Within schools, principals overwhelmingly said they focused new attention on the low-performing students, with that focus increasing from about 63 percent of the principals to over 80 percent in 1999. From responses to open ended questions, we learned that in some cases principals had access to additional funding that they could use to assist the low-performing students. In many cases, however, it appears that principals had either to shift resources away from the other groups of students or had to ask teachers to spend additional "voluntary" hours after school or on Saturdays working with these students. Among the other actions taken by various principals were increased use of tutoring, more individual assistance, more grouping of students by ability, greater use of computer technology (mentioned by only a few principals), and restructuring of the reading program. This greater focus by principals on low-performing students within schools is a desirable outcome that is fully consistent with the objectives of state policy makers.

Movement of teachers among schools. Across schools, however, the situation is more complex. One concern emerges from the observation that principals in schools with large proportions of low-performing students were less supportive of the ABCs program than were other principals and that they were less optimistic than other principals about their increased power to remove low-performing teachers. Such views could reduce the future willingness of ambitious and effective principals to serve in such schools. Second, the surveys suggest that schools with low performing students may find it even more difficult than in the past to retain or attract the higher quality teachers. While such teachers have always had incentives to move away from such schools to schools with more motivated and easier-to-teach students, the ABCs program enhances those incentives. By moving to "better" schools, teachers can now increase the chances of receiving a bonus and can minimize the chances of being associated with a publicly identified failing school. Importantly, however, because of its focus on gains in, rather than levels of, student test scrores, the ABCs program provides weaker incentives for teachers to leave schools serving disadvantaged students than would be the case if the program rewarded schools based on average levels of performance.

Empirical support for the possibility that schools serving low performing students might find it more difficult to retain teachers after the onset of the accountability system

emerges from a separate study of the movement of the movement of teachers among schools (Clotfelter. Ladd , Vigdor and Aliaga, 2004). In particular, we found that the schools that had disproportionate shares of low-performing students or that were officially labeled as "low-performing" schools found it harder to retain teachers after the introduction of the accountability system than before**.** This conclusion is based in part on simple comparisons of the trends in teacher retention rates in schools with large proportions of low-performing schools before and after the introduction of the ABCs program. That simple descriptive analysis is reinforced by a more sophisticated model explaining the probability that a teacher would leave a particular school in a particular year, having stayed in that school up to that point. The bottom line is clear. The schools with the lowest performing students had even more difficulty retaining teachers after the onset of the accountability program than before, and hence faced additional costs and challenges in their quest to provide a coherent and effective school program. Less clear, however is whether the introduction of the ABCs program actually reduced the quality of the teachers in the low-performing schools. The lack of clarity reflects in part the weak measures of quality we used in that study as well as the many other changes that were occurring in the North Carolina teacher labor market at that time, including, for example, the growing use of out-of-state teachers.

The general point here is that policy makers need to pay attention to the larger systemic effects of accountability system as well as their immediate effects in particular schools. To the extent that such programs provide incentives for greater turnover of teachers in schools serving low-performing students, an outcome that would be detrimental to the learning of such students, policy makers need to be vigilant about countering such incentives with specific policy actions designed to make such schools more attractive.

3. Design Matters – Subgroups under NCLB

Few would disagree with the conclusion that the design of a school-based accountability system matters. Already referred to are design issues related to how to measure the effectiveness of schools such as whether to use levels of performance, changes in average performance from one year to the next, or gains in the performance of specific cohorts of students..

Another central issue related to the design of accountability programs is the requirement under No Child Left Behind that schools be judged based not only on their overall performance, but also on the performance of each of 9 subgroups within the school. These subgroups are 1) White; 2) Black; 3) Hispanic; 4) Native American; 5) Asian/Pacific Islander; 7) Muliracial; 7) Economically Disadvantaged Students; 8) Limited English Proficient Students; and 9) Students with Disabilities. This subgroup requirement is intended to encourage schools to pay attention to the achievement of members of every subgroup and thereby to "leave no child behind." In the language of NCLB, the failure a school to demonstrate adequate yearly progress (AYP) for any one subgroup within a school means that the school has not been successful, where AYP is

the progress that would be needed for members of that subgroup to reach proficiency by the year 2013–2014..

The North Carolina State Board of Education has ruled that 40 students are needed for a subgroup to be large enough to be counted and has now incorporated the AYP reporting results in the ABCs program. For 2002-03, the AYP results are striking: more than half of all schools (53 percent) did not meet their AYP goals, presumably because of the failure of one or more of the subgroups within the school.

Table 1 presents the AYP results by ABCs recognition categories. At the top are the two categories of schools meeting their high growth or expected growth standards. Among the schools that are high growth under the ABCs program, 44 percent did not meet the AYP standards and among those that met their expected growth, 73 percent did not meet the AYP standards. Thus, there is a real discrepancy between the standards that the state has been using apparently quite successfully for several years and those required under the federal legislations.

The other ABCs designations are based primarily on a combination of levels of achievement and growth of achievement. Full definitions are provided at the bottom of the table. As indicated by the table, the proportion of schools within each category that do not meet their AYP standards increases monotonically across the school designations. . Only 21.6 percent of the Schools of Excellence (those in which over 90 percent of the students are at grade level) failed to meet the AYP standards, while almost 79 percent of the Schools of Progress (those in which 60 –80 percent of the students are at grade level) failed to meet the AYP standards. At the bottom, almost all of the Priority Schools and none of the 6 Low-Performing Schools met their AYP standards.

.

Evaluation of the subgroup requirement. These results bring to the fore some serious issues related to the desirability or lack thereof of the use of subgroup categories at the school level. On the one hand, one could interpret the North Carolina patterns as providing strong support for attention to the subgroup on the ground that past inattention to the subgroups has apparently allowed many children to be left behind. On the other hand, any system in which more than half the schools emerge as failing schools may well say more about the failings of the system than of the schools themselves.

The subgroup requirements of the federal NCLB Act notwithstanding, it is reasonable to step back and ask whether it makes sense to focus so much attention on subgroups at the school level. For a number of reasons highlighted most fully in research by Kane and Staiger (2002 and 2003), it probably does not.

First and foremost are the statistical problems associated with small numbers. Whenever there are small numbers of students, the chances are high that random variation in student test scores will outweigh any true signal about gains in test scores. This concern applies as well to the evaluation of small schools but is particularly serious with respect to subgroups where the numbers are frequently likely to be below 50.

Second, the fact that any State Board of Education is likely to set a floor for the size of subgroups, as NC did with its floor of 40 students within a school, sets up a dynamic that may not be desirable. It sets up strong incentives, for example, for schools to keep subgroups below 40 so that they fall below the radar screen. Further as shown by Kane and Staiger (2002), the requirement puts any school with multiple subgroups at a disadvantage relative to other schools and thereby discourages diverse student bodies.

Third, there is no compelling evidence that the subgroups generate the desired result of higher performance for those subgroups. A careful study of this issue of schools in Texas and California found no evidence that holding schools accountable for the performance of subgroups such as African Americans and Latinos had any effect on the performance of those students (Kane and Staiger, 2003).

The district as the more appropriate unit of accountability. To the extent that state policy makers wish to focus attention on the academic performance of subgroups of students, it would seem far more appropriate to do so at the district level rather than the school level. Current Federal requirements not withstanding, this line of argument suggests that state policy makers might do well to explore this alternative approach, perhaps with the goal of eventually overturning the federal legislation that is, arguably, inappropriately focused on individual schools.

Shifting the focus to the district level provides three major advantages over the school level. First, the numbers of students in the various subgroups are likely to be far larger which reduces the statistical problem of chance variation in the data. As a result, measured gains in student performance will provide a more accurate picture of true gains in performance. Second, this approach would be more consistent with the spirit of leaving no child behind in that the achievement of members of such subgroups who now happen to be in schools that fall below the threshold number could no longer be ignored. Third, and perhaps most important from a policy perspective, the district has a broader set of tools than do individual schools for addressing the challenge of improving the achievement of such students.

In addition to the ability to distribute resources among schools in ways to support the learning of students in lagging subgroups, district policy makers also have the power to affect how both teachers and students are distributed among schools within the district, and these distributions appear to matter for student performance. Research by one of my students at Duke University on the determinants of variation across districts in the performance of at-risk students indicates that one of the important variables is the evenness with which at-risk students are distributed among schools throughout the district. In particular, for any given percentage of at-risk students in a district, the less concentrated are the at-risk students in particular schools the higher appears to be the average performance of the at-risk students in the district. (Yankovich, 2003). Thus, for example, when districts such as Wake County, North Carolina, assign students to schools in such a way that no school ends up with more than 40 percent of its students eligible for subsidized lunches, the district is helping low-income students to achieve. Alternatively

when districts permit their low-income students to be concentrated in particular schools, they are making it more difficult for those students to learn.

Similarly districts play a large role in affecting how teachers are distributed among schools and that, too, has a potentially significant effect on the learning of subgroups of students. To the extent that district policies result in high teacher turnover and the placement of large numbers of novice teachers in the schools with the greatest concentrations of low-income, minority, or low-performing students, the students in those schools are at a disadvantage. The literature on teacher effects quite clearly shows that , regardless of how effective they may eventually become, novice teachers are less effective in generating student achievement than are more experienced teachers (Hanushek, Kain and Rivkin , Clotfelter, Ladd, and Vigdor , forthcoming.). Thus, if the ultimate goal is to leave no child behind, district policy makers will need to pay significant attention to how teachers, particularly novice teachers, are distributed among schools. The contention, here, is that districts would be forced to pay more attention to that issue if accountability for the performance of the various subgroups of students was directed at the district level rather than at the school level.

4. Accountability and Parental Choice

One of the provisions of NCLB is that once a Title 1 school fails to make adequate yearly progress for three years in a row, parents have the option of moving their child to another, more successful school. There are at least two reasons for including parental choice in an accountability program. One relates to fair treatment of students who ordinarily would have few options to leave a failing school. In the North Caroliina context, the large proportion of schools that are failing to meet the AYP standards raises concerns about the success of such a policy in meeting the fairness goal. In some cases there simply will be no available school within a reasonable distance. In addition, some of the receiving schools may have been able to meet the AYP standards only because they had too few students in some subgroup categories for those categories to count. To the extent that large numbers of members of lagging subgroups move from failing schools to such schools, the number of students in some of the subgroups might then pass the threshold level and have to be counted separately in the schools' performance. Thus, the movement of students among schools could put the receiving schools in jeopardy of failing to meet the federal subgroup standards.

The other reason often cited for including a choice component in an accountability system is to increase the incentives for the failing school to improve. According to this view, as long as a failing school has a captive group of students, it has little incentive to improve. Only when students have other options will the school face the necessary incentives raise student achievement. In early 2001, Jay Greene of the Manhattan Institute tested this proposition in the context of Florida's accountability and school voucher program. Under the Florida system, schools receive grades of A,B,C, D or F, but only the lowest performing schools (those rated F) are subject to the threat of a voucher. If a school received two Fs within four years, its student may receive vouchers,

which they can use to attend private schools. Greene compared the higher scoring F schools with the lower scoring D schools, under the assumption that the schools were quite similar in initial academic performance. Greene found that after receiving one failing grade the F schools showed greater improvement on standardized student tests than did the D schools. Noting that the schools were otherwise similar, he claimed in a widely disseminated report that the one significant difference between them – the threat of vouchers – accounted for the greater gains among the F schools.

A study based on North Carolina data, however, disputes that conclusion. Ladd and Glennie replicated the Florida analysis for North Carolina which does not have a voucher program. Treating the Low-Performing schools as comparable to Florida's F-rated schools and the No Recognition schools as comparable to Florida's D-rated schools. the authors found that even without the threat of a voucher, North Carolina's Low-Performing Schools exhibited greater gains in test scores than its No Recognition schools. Thus, they concluded that the success of Florida's F schools most likely had much more to do with the shame, attention, and additional assistance given to those as part of the state's accountability program than with the voucher threat.

Hence, policy makers in other states should be under no illusion that the choice provisions of the NCLB legislation will generate greater achievement in the schools to which they apply. Instead, the possibility that they will lead to an even greater concentration of disadvantaged students in such schools may well exacerbate the challenges facing those schools.

References.

Clotfelter, Charles, Helen Ladd and Jacob Vigdor. 2004. "Teacher Sorting, Teacher Shopping, and the Assessment of Teacher Effectiveness." Sanford Institute of Public Policy, Duke University.

Clotfelter, Charles, Helen Ladd, Jacob Vigdor and Roger Aliaga. 2004. Do School-Based Accountability Systems Make It More Difficult to Retain and Attract High Quality Teachers?" *Journal of Policy Analysis and Management.* Vol. 23, no. 2,.

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

Kane, Thomas and Douglas Staiger. 2002. "Volatility in School Test Scores: Implications for Test-Based Accountability Systems." *Brookings Papers on Education Policy 2002.* Washington, D.C.: Brookings. PP. 235-83.

Kane, Thomas and Douglas Staiger. 2003. "Unintended Consequences of Racial Subgroup Rules." In Paul E. Peterson and Martin West, eds, *No Child Left Behind? The*

Politics and Practice of School Accountabiity. Washington, D.C. : Brookings. Pp. 152-176.

Ladd, Helen and Glennie, E. 2001. "Claims for School Voucher Success in Florida not Justified." Policy Brief, Center for Child and Family Policy, Sanford Institute of Public Policy, Duke University.

Ladd, Helen F. and Arnaldo Zelli.2002. "School-Based Accountability in North Carolina: The Responses of School Principals." *Education Administration Quarterly.*

Ladd, Helen F. and R. Walsh. 2002. "Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right." *Economics of Education Review.*

Milgrom, Paul and John Roberts. *Economics, Organization, and Management.* Englewood Cliffs, NJ: Prentice Hall, 1992.

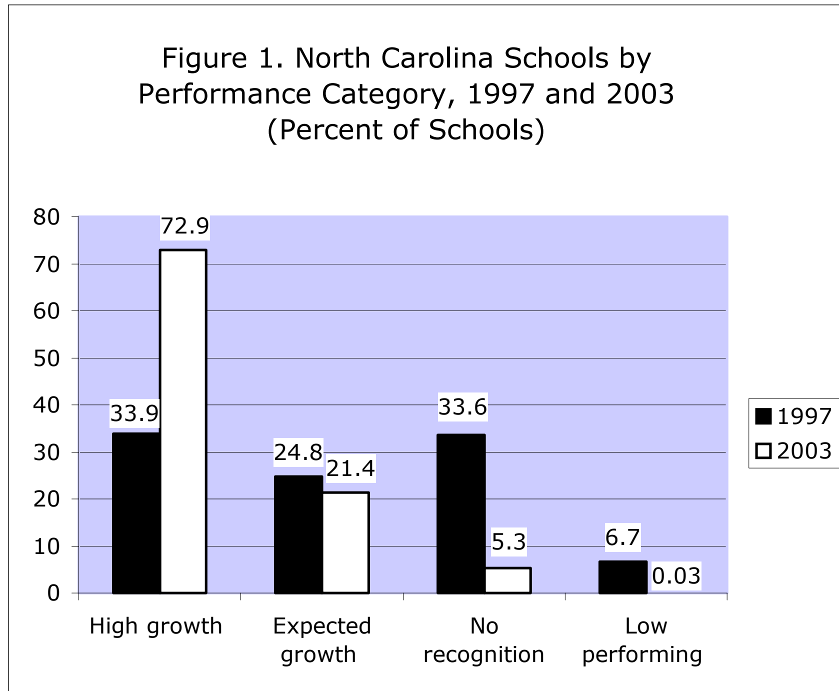
Yankovich, Michael. 2003. "Equal Opportunity for a Sound Basic Education." Policy Memorandum Prepared for the North Carolina Education Research Council. Master's Project, Sanford Instititute of Public Policy, Duke University.

Table 1. AYP Results by ABCs Recognition Categories, 2002-03

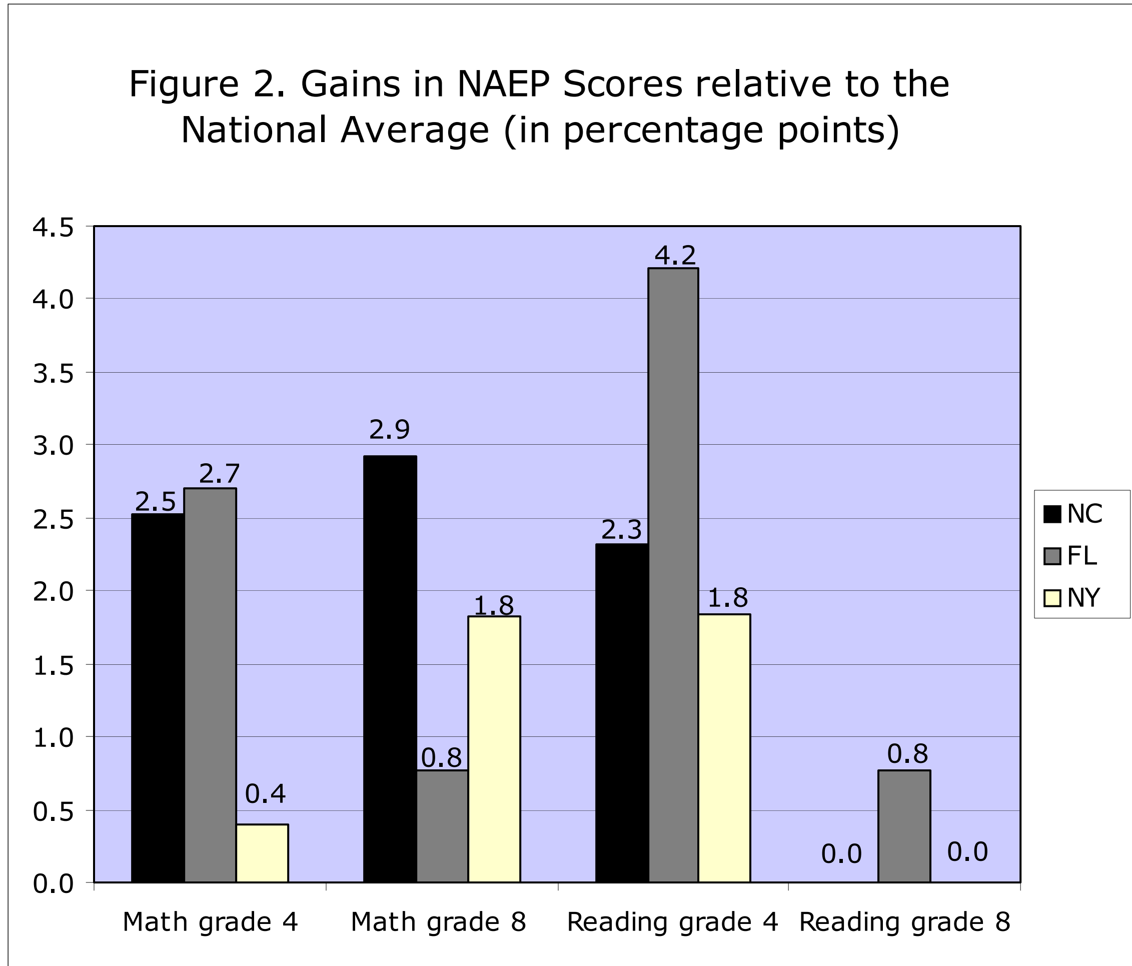
Category	**Met AYP**		**Did Not Meet AYP**		**Total**
	#	%	#	%	#
High growth	893	55.2	724	44.8	1,617
Expected growth	127	26.7	345	73.3	472
Schools of Excellence	371	78.4	102	21.6	473
Schools of Distinction	460	51.9	426	48.1	886
Schools of Progress	168	30.5	382	69.4	550
No Recognition	26	24.1	82	75.9	108
Priority Schools	2	2.6	76	97.4	78
Low-Performing Schools	0	0.0	6	100.0	6

Definitions. AYP stands for Adequate Yearly Progress as defined by the No Child Left Behind legislation. High growth and expected growth are the growth categories defined under the ABCs program (see text). Schools of Excellence have over 90 perecent of their students at grade level; Schools of Distinction have between 80 and 90 percent of their students at grade level, and Schools of Progress have between 60-80 percent of their students at grade level. Schools in all three of categories meet their growth expectation. No recognition schools did not meet their expected growth standards but have at least 60% of their students' scores at or above grade level. Priority Schools are those schools that have less than 60 percent of their students' scores at or above grade level, irrespective of making their expected growth standards, and are not Low-Performing Schools. Low-Performing Schools are those that failed to meet their expected growth standards and have less than 50 percent of their students at or above grade level.

Source. "The ABCs of Public Education: 2002-03 Growth and Performance of North Carolina Schools." (Last printed 10/13/2003.)



Notes. See text for description of performance categories. The No recognition category for this figure includes the 13 schools designated as priority schools that did not meet their expected growth.



Notes. The gains in math scores are for the period 1996-2003 and for reading scores are for 1998-2003.