# EDUCATION FINANCE RESEARCH CONSORTIUM

The Education Finance Research Consortium is an on-going collaborative research and policy analysis venture between the New York State Education Department, the New York State Board of Regents, and education scholars, researchers and policymakers from New York and around the country. By bringing education researchers together with policymakers, the Consortium seeks to address critical questions about the state of public education in New York. The Consortium's mission is to foster an exchange of information and ideas that informs education policy decisions, assist education policy decision-makers in identifying issues that require initial or further study, and produce quality and timely research that contributes to the development of sound education policy. The Consortium is housed in the Center for Policy Research, Rockefeller College of Public Affairs & Policy, University at Albany – SUNY.

Condition reports are written using education data compiled by New York State agencies and from data independently collected by the researchers. Consortium researchers work closely with agency staff to identify and analyze trends in school district spending, school staffing and student performance. The condition reports describe various characteristics of New York State's public school systems and also identify potential areas for further study.

The Consortium's Governing Board would like to thank the New York State Education Department and the New York State Board of Regents for their intellectual and financial support. We are particularly indebted to Commissioner Richard Mills and Deputy Commissioner James Kadamus. Deborah Cunningham of the New York State Education Department is liaison between the Consortium, researchers and Department policy makers, provides feedback on draft reports and reviews research findings for their relevance to Regents' policies. Barbara Downs of the Center for Policy Research works with researchers to coordinate the presentation of their findings at public forums and symposia, and to publish Condition reports.

This Condition Report is one in a series of published Consortium reports. To access other Consortium publications or to get information on Consortium-sponsored research, please visit the Consortium's website at http//:www.albany.edu/edfin, phone (518) 442-3862, or email edfin@albany.edu.

# Uncommon Schools, Uncommon Results:
# Case Studies of Three New York State Schools Closing Racial Test Score Gaps

Colin C. Chellman

Meryle Weinstein

Leanna Stiefel

Amy Ellen Schwartz

New York University

December 2004

***Note:*** Readers are directed to the Education Finance Research Consortium's website at http://www.albany.edu/edfin for the Appendices to this condition report. The appendices include important information related to the authors' study.

## Executive Summary

This paper examines schools with atypical racial test score gaps, where non-white students have "beaten the odds" by performing as well as or better than white students. Using a unique dataset of student ELA (English Language Arts) and math pass rates for all schools in New York State, we identify these atypical schools and describe them statistically. We then choose three schools for study using qualitative methods of inquiry that work with hypotheses from other qualitative studies of school-level test score gaps.

### Selected Findings

Findings of the qualitative work suggest a number of factors that may contribute to school success in eliminating gaps:

- on-going assessment and data tools -- as well as professional development training on using data tools -- to target students for special attention and resources;
- teachers' understanding of data and their ability to link data to instructional practices;
- strength of principal's leadership skills;
- principals with direct involvement in hiring most of their faculty, making them more trusted and respected by their teachers, and in turn, allowing them to trust their teachers to effectively address student needs;
- high parental expectations; and,
- high expectations for *all* students communicated widely by teachers and principals.

### Recommendations

Based on these preliminary findings, we recommend additional research that includes a comparison group of schools that are not successful in reducing test score gaps.

## I. Introduction

The existence and persistence of a gap in test scores among students of different gender, racial and economic backgrounds is well-documented, both for the country as a whole and for New York State in particular. As shown in Stiefel, Schwartz and Chellman (2003), however, while most New York State public schools show the familiar test score gaps, some schools do not follow this pattern. Instead, these schools show comparable performance across subgroups or, in some cases, show "reverse gaps," in which a typically (on average) low-scoring group outperforms a typically high-scoring group. This paper builds on these results to better understand the characteristics of the schools that have " beaten the odds" by reversing persistent and familiar achievement patterns.

The paper's quantitative analyses build directly upon the findings of the earlier report (Stiefel *et. al.*, 2003), using 4th and 8th grade math and ELA test scores for every school in the state for school years 2000-2001 and 2001-2002. These data include average test scores for all students in each school, as well as average scores for five different student subgroups defined by race/ethnicity, gender, economic status, English language proficiency, and migrant status (New York State Department of Education, 2002). These detailed data allow subgroup analyses of test scores at the school-level, as well as comparative analyses across time. We focus on race in this paper.

The qualitative analyses focus on a set of schools that consistently show comparable or atypical performance across subgroups. We carried out a series of case studies to identify the school and district policies, curricular programs, and instructional practices of schools that help these atypical schools to succeed where other schools do not.

This paper investigates the following questions:

1. What are the differences in school-level test scores by subgroup in New York State in 2001-2002? How does the pattern of school-level test scores observed in 2001-2002 compare to the pattern in 2000-2001?
2. How do the schools that consistently show comparable performance across subgroups -- the "schools without gaps" -- differ from those with consistent disparities in performance across subgroups?
3. What are the district and school policies, as well as teacher and principal practices and beliefs, that might contribute to reducing test score gaps?

We pursue the first two of these questions quantitatively, the last one primarily qualitatively. We use quantitative methods to analyze and validate the nature and extent of the effectiveness of these atypical schools in bridging the achievement gap; we use qualitative methods, particularly school assessment instruments, to identify and analyze the key organizational and instructional practices critical to successful outcomes in a small sample of schools. These methods are triangulated to form a profile of New York State's schools that have beaten the odds in bridging the achievement gap.

This paper is organized into four sections. We begin with a two-part literature review that first examines hypotheses about why racial test score gaps exist, then reviews strategies proposed to eliminate test score gaps between race groups. The second section presents data sources and the methods used to examine a variety of strategies schools could use to eliminate test score gaps. In the third section, we present findings, and in the fourth and final section, our conclusions.

## II. Literature Review

There is considerable evidence of the persistence of gaps in academic performance between white and non-white students (specifically, Hispanic and black students) in both state and national data on standardized test scores from early childhood to high school. For example, Hedges and Nowell, in Jencks and Phillips' *The Black-White Test Score Gap* (1998) as well as a follow-up article (1999), use high school test score data from six nationally representative surveys conducted between 1965-92 to show that the test score gaps between whites and blacks declined over this period, that adjusting for socioeconomic factors reduces gaps by less than half, and that the trends in unadjusted and adjusted test score gaps follow similar paths. Cook and Evans (2000) use data on middle school students from NAEP between 1970-88 to decompose reductions in the black-white test score gap; they find that 25 percent of the change can be attributed to family and school characteristics, while 75 percent can be attributed to within-school reductions. More recently, Fryer and Levitt (2004), using the kindergarten through first grade cohort data from the Early Childhood Longitudinal Study (ECLS), find that adjusting for background characteristics of students reduces the black-white achievement gap to zero at kindergarten entry (but not thereafter).

Although the existence of achievement gaps is supported by empirical evidence, the causes of these gaps are less well understood. We review prominent hypotheses about the reasons for the existence of test score gaps, and then review proposed strategies for eliminating test score gaps. From this examination of the literature, we shaped the questions we included in our surveys and interviews.

### Why Do Racial Test Score Gaps Exist?

We focus on hypotheses that explain raw, "unadjusted" test score gaps. Even after controlling for poverty and other factors that may play a role in producing test score gaps, these gaps in test scores between race groups remain. Note that only one study (Fryer and Levitt, 2004) has found that gaps disappear after adjusting for home and neighborhood factors, and thus no hypotheses are completely satisfactory explanations for the existence of test score gaps.

**Home factors.** Two of the most frequently cited home factors implicated in test score gaps are family poverty status and the educational attainment of parents, both of which strongly correlate with race. There could be a number of reasons for this effect of poverty, such as poor nutrition, lower birth weights, and fewer educational resources available in the home. Thernstrom and Thernstrom (2003) suggest that the effects of poverty -- lack of books in the home, for example -- exacerbate the limited schooling effort of students of color, thus contributing to the achievement gap. It is also possible that student achievement is affected by

factors that are closely associated with poverty and level of parental education, such as family wealth or assets and quality of schools attended by parents, grandparents, or other relatives with whom children have contact.

There is some circularity concerning the problems of poverty, race, and education. Poverty is correlated with race. Problems associated with poverty enter the classroom (in the form of more special needs students and more students needing social services, etc.[1]), and if it is more difficult to educate children in areas with higher levels of poverty, then it is less likely children will acquire the skills they need to get jobs and escape poverty.[2] Since segregation by income and race are also correlated -- particularly in urban areas and for African-Americans -- there will be disparities between the education services students receive in urban areas versus what students receive in suburban areas.[3] These disparities in services will lead to disparities in outcomes. While researchers such as Huggins and Celio (2002) and Rothstein (2004) suggest that programs to reduce poverty might help to reduce the achievement gap, those policy options are not available to schools.

**School factors.** Ellen et. al. (2002) describe three key ways schools (and, in particular, school segregation) may affect student achievement and employment outcomes: 1) peer effects, 2) differential resources; and 3) network effects. The peer effects hypothesis contends that decisions about study habits, standards of behavior, and the kinds of classes a student takes are shaped as much by peers as by parents. There is consensus on these effects except for high-achieving students (Ellen et. al., 2002). Some researchers have hypothesized that schools with test score gaps might engender cultures that inhibit students of color from taking more challenging classes. Ferguson (2001), in his study of students in Shaker Heights, OH, found that African-American students were greatly under-represented in honors and Advanced Placement classes, and hypothesized that one reason for this under-representation might be students' fear of social isolation as a member of a very small minority group in those classes.

Allocations of resources within districts reflect politics, differences in costs, and the decisions of teachers, principals, parents, and students over time. There is an extensive literature on whether resources -- and which resources -- might affect student achievement and other student outcomes, with not much consensus about the answer. One of the earliest studies by Coleman *et. al.* (1966) states that family background is more important than school inputs in the determination of achievement. Since then, prominent researchers such as Hanushek (1986; 1997), Greenwald *et. al.* (1996), and Hedges *et. al.* (1994) have come to

[1] Orfield (2001) explores reasons for lower levels of educational achievement in urban schools. He explains that one reason is the level of poverty among families served by those urban schools; schools with large numbers of high-poverty students necessarily will devote more time and resources to social services to deal with family and health crises, school security, students who do not speak standard English, and students who are poorly prepared for education.

[2] For example, Freeman and Holzer (1986) studied the employment outcomes for African-American youth in central cities and found that increases in both the quantity of education (measured as time spent in school) and the quality of education (measured by grades received by the students) increased wages and the number of hours worked.

[3] Schiller (1989), for example, finds that the high school education received by urban African-American students is roughly equivalent to the education received by a white high school dropout.

conflicting conclusions.[4] It is generally acknowledged that schools differ in quality, but measuring quality is difficult. Findings from numerous studies show that total and instructional expenditures, class size, or other commonly measured school attributes are not significant factors in student achievement using currently available data and methods. Card and Krueger (1996) find that increased investment in school "quality" among states is consistently associated with higher earnings of adults who were schooled in those states, holding constant other influences.

This point aside, from an equity-of-inputs standpoint, with whom one goes to school might affect levels of inputs/resources (Stiefel *et. al.*, 2005, in press). There is evidence that schools in districts with more non-white students receive fewer resources in New York State (e.g., again, see Stiefel *et. al.*, 2005, in press).

In addition, it is possible that the people with whom one attends school might have effects other than on allocations of resources. There may be differential longer-term effects of "productive" versus "nonproductive" networks. Wells and Crain (1994) found that African-Americans who attended desegregated schools benefited over the long-term because of superior connections to the "white world," but they also found few short-term benefits as measured by test scores. Orfield (2001) found that there is a strong correlation between the percent poor in a school and its average test scores, and concluded that non-white students in segregated schools, no matter how able they may be as individuals, usually face less competition and a much lower level of preparation by other students.

**Societal factors.** Although it is difficult to quantify, discrimination may play a role in the type and quantity of education received by students of color (Huggins and Celio, 2002). This belief can also affect the way students of color think about themselves, leading to lower performance, or lower expectations for themselves. Steele and Aronson (1998), for example, believe that stereotype threat may affect the performance of students of color, as they may have anxiety about conforming to stereotypes of non-white performance on standardized exams. Unfortunately, the number of hypotheses concerning societal factors' role in producing underachievement is extensive. Since we are unable to test any of these hypotheses, we leave this vast literature for future study.

## Strategies to Eliminate Test Score Gaps

Theories about the causes of test score gaps lead inexorably to strategies to reduce these gaps. We limit our review to factors under the control of schools, the unit of analysis for this study. Some strategies are

[4] Burtless (1996) summarizes much of this literature. Card and Krueger (1996), however, note that some studies suggest that the relationship between educational spending and wages is weak or nonexistent, so the issue has not been resolved. In Hanushek's (1997) meta-analysis of education production function studies, he states that teacher skills -- based on teacher IQ or achievement tests, but not level of education, years of experience, or salary -- appear to be most important to improving student outcomes, while noting that increases in costs and quality of the inputs (using common measures of quality) appear unmatched by student performance. He notes that commonly used measures are generally proxies since many outcomes cannot be quantified. It is generally accepted that schooling does matter since there is high positive correlation between higher levels of schooling and positive attributes after schooling. There is a clear need to measure differences in quality and quantity rather than treating all schooling as equal, for example most education production function literature does not include a measure of the value of time spent in different settings. The inclusion of family background variables also generally seems to show that family background correlates with school resources and has strong effects on student outcomes (Hanushek, 1997). The ideal analysis would match individual students with school and family resources.

focused more on low-income students than on non-white students, but given the high correlation between race and income, we include them.[5]

## School and Teacher Culture and Characteristics

**Small Class Size and High Per-Pupil Spending.** Class size and classroom structure are two factors that might be manipulated to help low-performing students succeed, thus reducing test score gaps. For both strategies there is mixed evidence, although the former has experimental research evidence that provides some support for its efficacy in raising low-performing students' test scores, although the effects were limited and did not grow over time, as would be expected.[6] Other studies using econometric techniques to examine the effects of class size, such as Rivkin, Hanushek, and Kain (1998), were able to find statistically-significant but small effects of class size differences, with effects sizes differing by socioeconomic status of the student's family.

On a broader level, small class sizes require higher levels of per pupil spending. Card and Krueger (1996) summarize the results of several studies that examine the relationship between school spending, or resources, and earnings. There is evidence that school resources matter: students who were educated in schools with relatively high spending levels earned higher wages. As Card and Krueger note, some studies suggest that the relationship between educational spending and wages is weak or nonexistent, so the issue has not been resolved. The literature on the effectiveness and efficiency of school spending highlights the need to develop models that link education to a larger range of outcomes, such as earnings, health status, public assistance, crime, reduced training costs, among others.[7]

**Teachers.** Despite the controversy over the link between resources and achievement, there is compelling evidence that if those resources come in the form of high quality teachers, then resources lead to better student outcomes.[8] For example, Jepsen and Rivkin (2002) examined California's class size reduction efforts and

[5] While researchers have suggested that income redistribution programs may lead to better educational outcomes for poor students, those policy options are not available to schools or districts. For example, schools are unable to directly reduce levels of poverty in their broader community, but they can be, and are, often charged with reducing the effects of poverty.

[6] The Tennessee STAR Experiment attempted to discern the effects of specific resources on student performance by using a random assignment experimental design as an appealing alternative to more commonly used method of econometric estimation. Such an approach can help to overcome a variety of concerns raised about selection and causation. A group of kindergarten through third graders in Tennessee was randomly assigned to either large classes (22-24 students, with or without a teacher aide) or small classes (14-16 students). The student testing showed that children in smaller classes did better at the end of kindergarten and that this better performance was maintained through the third grade (Word *et. al*., 1990). However, the actual implementation of this experiment is open to question, especially concerning the randomization procedure, with significant numbers of students moving from large to small classes. The experiment's analyses yielded results that were small and difficult to interpret. The initial achievement differences found in the year students entered a small class were maintained but did not become wider through the grades (Word *et. al*., 1990; Krueger, 1999). Because they continue to get more resources (smaller classes), these resources should, according to the general hypothesis, keep producing increased student performance.

[7] General conclusions about the effectiveness and efficiency of school and teacher resources were summarized by Hanushek (1996; 1997; 2002). Schools differ in quality, but Hanushek maintains he is not certain what quality is, since findings of many studies show that expenditures, class size, or other commonly measured attributes are not significant to student outcomes. Teacher skills appear to be most important, though, in his review of the literature. Increases in costs and the quality of the inputs appear unmatched by student performance, although he notes that commonly used measures are generally proxies since many outcomes cannot be quantified.

[8] A study of inner-city schools found that during a single academic year, a student with a high-quality teacher outperforms a child with a low-quality teacher by up to one full grade level (Hanushek 1992). While researchers agree that some teachers are more effective than others, they do not agree on what makes a good teacher. In looking for teacher characteristics that explain differences in

found that there was a trade off between teacher quality and class size, especially for African-American students. Snipes *et. al.* (2002), in their district-level study for the Council of the Great City Schools, documented the challenge of teacher inexperience and turnover to efforts to improve poorly-performing students' achievement, and highlighted the challenge of low teacher expectations to efforts to help poorly-performing students achieve greater academic success.

Not only is there empirical support for reducing teacher turnover and improving teacher quality, there is also policy support. Most states have initiatives to attract and/or retain good teachers to improve the performance of their lowest-performing students. There are many examples, such as the Connecticut's Commissioner of Education's Back to School Address of August 18, 2004, in which she announced proposals to reduce test score gaps by retaining qualified teachers in urban areas through performance bonuses and multiyear contracts, as well as proposals to bring teachers out of retirement to work in urban school districts (Sternberg, 2004, p. 10).

**Integration and Student Relationships.** As addressed earlier, there is an extensive literature addressing the benefits of integration as well as the negative consequences of segregation. One possible consequence of segregation is disparities in the provision of public services. These disparities occur because of the structure of metropolitan areas, where local service-providing districts -- which, as discussed earlier, are often divided by race and income -- depend primarily on their local tax bases for financing.

Disparities in public services, such as education and public health, may be translated into disparities in outcomes among residents of jurisdictions in terms of future income and employment. For example, Ellen (1999) explored the link between racial segregation and city-suburban disparities (using an index of those disparities, including three outcomes measures: per capita income, employment rate for working age men, and the proportion of people 25 or older who have at least a high school diploma). After controlling for region, population, poverty rate, and the percentage of African-Americans and Hispanics in the metropolitan area, Ellen found that segregation (both white/African-American and white/Hispanic) is positively related to disparities between the central city and suburbs. These consequences of racial segregation may be exacerbated by income segregation.

Segregation may also have more direct consequences for students through peer or network effects (as described, for example, in Ellen *et. al.*, 2002). Hanushek *et. al.* (2001) state that in most studies, a lack of interracial contact in elementary and secondary school can be harmful (vis-à-vis test scores, years of education, and earnings) to black children. An important exception is Rivkin (2000) who found little effect of

---

effectiveness, researchers have focused on education level (years of postgraduate study), communication skills (verbal ability), and experience (years of teaching). The results of the studies seem to show:
1) Education level: There is no evidence that teachers increase their teaching skills by going to graduate school. It appears that students do not learn more from teachers with advanced degrees.
2) Verbal skills: The most effective teachers have superior communication skills. Students learn more from teachers who score high on standard tests of verbal ability.
3) Experience: There is some disagreement about the effects of experience on teaching skills. Murnane (1983) suggests that teaching effectiveness increases for the first few years of teaching and then levels off. In contrast, Hanushek (1971; 1981) suggests that teaching skills are independent of teaching experience.

exposure to white students on black students (again, vis-à-vis scores, years of education, and earnings), and ultimately recommended the improvement of school quality over desegregation to achieve better test score and improved labor market outcomes for African-Americans.

**High Teacher Expectations.** Huggins and Celio (2002) note the following conclusions from the research on teacher expectations: 1) Teachers have lower expectations for black students; 2) teacher expectations have a disproportionate effect on black students; and, 3) teachers' expectations are based on statistical discrimination, where past performance and group averages are consciously or subconsciously attributed to current individual students, thus helping to perpetuate performance disparities.

Rothman (2001-02) notes that teachers in Department of Defense (DoD) schools have very high expectations of their black and Hispanic students, resulting in DoD schools that exhibit lower test score gaps than other schools. Ferguson (1998a) states:

> No matter what material resources are available, no matter what strategies school districts use to allocate children to schools, and no matter how children are grouped for instruction, schoolchildren spend their days in social interaction with teachers and other students. As students and teachers immerse themselves in the routines of schooling, perceptions and expectations both reflect and determine the goals that they set for achievement; the strategies they use to pursue the goals; the skills, energy, and other resources they use to implement the strategies; and the rewards they expect from making the effort. These should affect standardized scores as well as other measures of achievement. (p.274)

**Diverse Faculty and Staff.** Diverse faculty and staff may help to engage students of color, relate better to those students, or serve as role models, thus improving education outcomes for non-white students. Siddle Walker (2001) focuses her research on teacher characteristics; she enumerates a set of principles that seem to capture the beliefs of African-American teachers in particular held about their teaching role: 1) strong teacher identification with the culture and beliefs of black communities; 2) active involvement with those communities; 3) commitment to professional ideals/dedication to teaching and strong ethic of caring for children; and, 4) adaptation of curriculum to meet student needs and supplementation of the required curricula with information about the history, accomplishments and dilemmas of African-Americans in the US.

**Leadership.** School leadership is another area in which research and policymakers have focused attention to reduce test score gaps. The Connecticut Commissioner of Education, for example, is going to propose that promising students in school administration programs receive tuition reimbursement in return for working in targeted urban districts in the state (Sternberg, 2004: 10). In their well-known study of school organization and leadership, Chubb and Moe (1990) find that differences in school-level effects are largely due to the internal organization of schools; they argue that public schools are bureaucratic and encumbered by an institutional framework governed by direct democratic control. They conclude that private schools are more effective than public schools because private schools are similar to market participants, possessing autonomy, clear missions, strong leadership, and worker professionalism. Berliner and Biddle (1995) concluded that

teacher and administrator autonomy was the most important influence on student achievement. The hypothesis that principal autonomy and strong leadership help teachers and students excel is compelling.

**Strong Recruiting and Retention of Good Teachers.** Huggins and Celio (2002) point out that low-income students and students of color experience higher rates of teacher turnover than other students. They thus posit that schools with less turnover -- in particular, less turnover of good teachers -- will experience greater student outcomes due to improved staff morale, relationships with parents, and in-classroom performance.

## Use of Data in Schools

Perhaps the first step to addressing the problem of achievement gaps is to identify the problem, even though principals and teachers may be reluctant to do so for a variety of ideological reasons (e.g., fear of implicitly supporting arguments that gaps are due to innate ability, or fear that poor student performance will reflect badly on principals and teachers). Often, however, states and districts provide schools with data aggregated at various levels that do not provide information that is useful for teachers in their day-to-day work. Disaggregating student performance averages allows teachers to identify low-performing students. By identifying students who need the most help, teachers can intervene early in a student's education (Huggins and Celio, 2002). If data are disaggregated, useful benchmarks might not be provided to teachers and school administrators so that they could compare their own students' performance against some average across the state or district. And the provision of disaggregated data in a more useful format to teachers and school administrators does not guarantee teachers have the training to analyze the data and translate the information into effective instructional practices.

Recently, the Bay Area School Reform Collaborative (BASRC) undertook a series of case studies of select California schools; some of those schools had reduced racial achievement gaps and some had not (Symonds, 2003). BASRC found that collecting, organizing, and presenting data to teachers and principals was a key component in helping schools succeed. Prior to the BASRC study, Snipes *et. al.* (2002) carried out a district-level study for the Council of the Great City Schools that highlighted the challenge of low teacher expectations to helping poorly-performing students achieve greater success. They found that successful districts focused on low-performing schools and used data to inform decisions about instruction. These data were augmented by ongoing assessments to identify student and teacher weaknesses. In addition, the NEA Foundation released a report espousing the usefulness of data in identifying students who need help to improve; the report focused on the importance of providing professional development to support the use of such data (NEA Foundation, 2003).

This strategy has been undertaken by a number of states to improve outcomes for low-performing students and to reduce test score gaps. According to a recent NCES publication (2003), as early as the year 2000, four states were focusing on the use of data to analyze achievement gaps between subgroups of students as one of their primary methods of addressing their states' achievement gaps. In addition, the majority of states (27) by 2000 were using student test scores to evaluate schools. Because of the passage of NCLB in January 2001, the number of states pursuing this strategy has increased. For example:

- Calhoun *et. al.'s* (2004) presentation at the Alaska Department of Education NCLB Winter Conference focused on the importance of monitoring student performance to increasing test scores, and ultimately closing the test score gap by raising the performance of the lowest-performing students.
- The California State Education Superintendent issued a press release supporting the importance of data in reducing test score gaps in California -- the press release was prompted by the Superintendent's opposition to Proposition 54, which sought to eliminate data collection disaggregated by race or ethnicity (O'Connell, 2003).
- The Colorado Department of Education created a task force to reduce the achievement gap between white and non-white students. Two ways the task force has approached the reduction of the test score gap entail using data to promote accountability and promoting parent, family and community involvement in schools (CDE, 2003).
- In announcing reductions in test score gaps between white and non-white students, Georgia State Superintendent of Schools Kathy Cox attributed gap reduction to data disaggregated by race group, which allowed school leaders and teachers to target students who needed extra help (Cox, 2003).

## Professional Development and Instruction

**Ongoing assessment, translating data into instructional practice, focusing on low-performing students.** As highlighted by a recent NEA Foundation study (2003), using data alone cannot reduce test score gaps or improve the performance of the lowest-performing students. Professional development for teachers in interpreting data, translating data into effective instructional strategies, and then targeting low-performing students for specific instruction are all characteristics of professional development programs in successful schools identified in the literature. One of the key findings in Symonds' (2003) case studies of successful schools (and a comparison group of schools that were not successful in reducing test score gaps) was that teachers in successful schools were more likely to have received professional development about the use of data, worked under school leadership that encouraged the use of data, and discussed data and appropriate instructional strategies with colleagues. Huggins and Celio (2002), in their case studies of successful schools in Washington State, found that teachers in successful schools were trained to use data to target low-performing students.

The literature also suggests that teachers should receive training not only in interpreting data from state exams, but also from their own ongoing assessments. Rothman (2001-02), for example, reports that two of the common characteristics of schools successful at educating poor, non-white students are frequent assessments (i.e., multiple opportunities to succeed) and written responses to students tested on those assessments. These ongoing assessments also allow teachers to change instructional practices to meet the needs of low-performing students.

**Literacy Instruction.** Once teachers collect and interpret data, they must address the student problems that emerge from the analysis through appropriate instructional strategies. One important strategy suggested by the literature as particularly effective in increasing the performance of low-performing students and decreasing test score gaps is literacy instruction. Again, one of Symonds (2003) key findings for decreasing test score gaps was the implementation of, and focus on, literacy skills. Symonds found that schools were pulled in many directions instructionally, but those that focused on basic literacy skills fared the best at reducing test score gaps. A number of states are using this strategy to reduce test score gaps. Vaughan (2001) describes the initiative undertaken by Alabama to help 100 percent of the state's children achieve literacy. Connecticut emphasizes early childhood education -- in particular, literacy training -- to reduce test score gaps between white and non-white students (Connecticut State Board of Education, 2003). In 1998, the State of Florida entered into an agreement with the US Office of Civil Rights to improve access to education for minority students; part of the goal was to decrease test score gaps between minority and non-minority students. Two strategies highlighted for decreasing test score gaps over the five-year period included using data to decrease gaps and implementation of professional development in reading instruction for teachers (State of Florida, 2003).

**Parents and Expectations.** Parental expectations and involvement in their children's schools are topics mentioned in the literature as possibly having an effect on student achievement. Lareau and Horvat's (1999) case studies of parental involvement in their children's schools highlighted black parent concerns about discrimination in school. These concerns affected how black parents approached school faculty and staff, and effectively raised barriers to communication and potential improvement in their children's education. Lareau and Horvat's work raises important concerns about the ways schools try to engage parents, as well as how frequently that contact occurs. Despite these concerns, some states have made parental involvement a key component of their plans to reduce test score gaps. Colorado, for example, has identified parental involvement as an important component in their plans to reduce the state's test score gaps. Colorado's Department of Education created a task force to reduce the achievement gap between white and non-white students; the task force has advocated for reducing test score gap by using data to promote accountability and promoting parent, family and community involvement in schools (CDE, 2003).

Perry *et. al.* (2003) suggest that one way to help poor, non-white students succeed is by school faculty and staff regularly communicating expectations -- both in public or group settings as well as in individual interactions -- that every student perform at a high level. Further, Perry states that "…institutions that are culturally responsive and that systematically affirm, draw on, and use cultural formations of African-Americans will produce exceptional academic results from African-American students." (p.107) Perry then theorizes about what does not work in raising achievement levels for African-American students. She believes that programs and schools that are highly stratified, that focus on individual achievement and competition, and allow students a great deal of freedom to achieve, will fail African-American students. She supports

"leveling cultures" to help African-American students achieve -- i.e., no differing levels of classroom infrastructure, no ability grouping (i.e., math and reading only), and no tracking (all subjects).[9]

Although the literature provides no definitive evidence that any one strategy will eliminate test score gaps within schools or across districts or states, some strategies have received consistent policy and empirical support. There is evidence of the positive effects of high teacher, parent, and principal expectations for all students. These expectations should be communicated often and in public to everyone in the school community. Schools enjoying high quality teachers and little turnover, and strong leaders who allow teachers the flexibility to address their students' specific needs, also appear to be an effective strategies to reduce test score gaps. School leaders also emphasize the use of student performance data to identify students who need additional help, and provide resources to teachers to interpret the data and develop effective instructional strategies. In addition, a diverse student body that interacts and has relationships across race and language groups provides support for the improvement of education for all students. Although parental involvement in general is seen as a positive influence on a student's education, the most important form of parental involvement seems to be in the communication of high expectations. These strategies will be tested in our case studies of three New York State schools successful in reducing test score gaps.

## III. Data Sources and Methods

Using hypotheses and strategies suggested by the literature, we constructed surveys and interview protocols to collect information on three successful schools in New York State. The New York State Education Department (SED) provided the test score and school characteristics data that allowed us to select schools for the qualitative case studies. Specifically, the SED provided the New York State School Report Card 2001 and 2002, which includes school test scores by subgroup for school years 2000-01 and 2001-02, the Institutional Master File (IMF) for 2000-01 and 2001-02, which contain school and teacher characteristics, and the School District Fiscal Profiles for 2000-01 and 2001-02, which supply district-level finance data.

Qualitative data were collected from primary sources by the New York University team in two stages: first, principal interviews, and second, teacher surveys. The principal interview protocols were developed based on topics the literature suggested have an impact on student learning and reducing test score gaps: the school and principal's educational philosophy, academic goals for the school year, level of standardization or centralization of curricula across classroom, level of expectations of student performance, instructional methods, school supports, staff qualifications (leadership, instructional, and support staff), school environment (building and classroom), and school organizational structure. Teacher surveys probed for similar information. We asked teachers about their understanding of the school philosophy and academic goals for the year, as well as their perceptions of other teachers, parents, and students. Through the surveys,

[9] Addressing tracking and ability grouping, Ferguson (1998b) finds that accelerated classes increase gaps between high and low scorers since they benefit students who already score high. In other words, an ineffective method of decreasing the test score gap is to bring the high scorers' scores down. Oakes (1985) finds that teachers of higher track classes tend to be more enthusiastic and work harder.

teachers commented on the level of standardization of curricula across classrooms at same-grade level as well as the ability to be flexible to meet student needs. Surveys also addressed teachers' expectations and perceptions of students and student interactions (in and out of the classroom), perceptions about school leadership quality and priorities, ability of teachers to overcome obstacles to reducing test score gaps, and availability of teacher supports in the school.

### Quantitative Analysis

We identified the set of schools in which reliable subgroup pass rates could be calculated, and then created school-level variables to capture the disparity in performance between racial subgroups. Using these data, we replicated key analyses from Stiefel *et. al.* (2003) to analyze the distribution of schools and students across the state according to whether the school would be held accountable for white students only (i.e., too few non-white students tested to be accountable for non-white students as well as white students), non-white students only, or both groups of students. We then compared the results between two years (school years 2000-01 and 2001-02), explored differences and similarities, and identified a set of schools that showed little disparity in performance between subgroups for both years. A set of these schools-without-gaps were studied more fully using qualitative methods.

Reflecting the understanding that performance measures are volatile and unreliable when based upon small sample sizes, NCLB legislation made clear that schools with small numbers of students in a subgroup would not be held accountable for the subgroup performance. However, no specific threshold number of students was specified. In principle, the threshold should be high enough that the statistics are reliable and valid, but not so high that few schools will meet the threshold. If the standard is set too high, too many schools will be exempt or 'not accountable,' even if the statistical properties of the measures are advantageous. If the standard is set too low, the performance measures will be meaningless, even if many schools are 'accountable.' For this study, we set a threshold requiring 10 or more test scores in a sub-group (white or non-white), because current plans indicate that New York State will use between 20 and 40 students (an average of 30) across three grades, or an average of 10 students per grade.

Using this threshold, Table 1 shows the distribution of schools and students tested by racial accountability status -- whether a school will be held accountable for the performance of both white and non-white students. For elementary schools, 42.4 percent -- representing 34.6 percent of fourth grade students -- would be held accountable for white students only, while 23.3 percent of schools (representing 26.5 percent of fourth graders) would be accountable for non-white students only. That leaves 34.4 of schools (with 39 percent of students) that would be held accountable for both white and non-white scores. Elementary schools accountable for both groups, on average, are larger.

Similar trends are found for middle schools: 36.7 percent of schools -- representing 20.9 percent of eighth grade students -- would be held accountable for white students only, while 18.1 percent of schools -- educating 17.7 percent of non-white students -- would be held accountable for non-white students only. The remaining 45.2 percent of schools -- educating 61.4 percent of eighth grade students -- would be held

accountable for both white and non-white students. Again, middle schools accountable for both groups are larger on average.

A little more than a third of the elementary schools in the state -- and less than half the middle schools -- would be accountable for the performance of both white and non-white students; these findings could reflect the legacy of residential segregation (Orfield and Lee, 2004). New York is a state with high levels of local control, and based on the literature, one could assume a great deal of sorting by race and income group across districts. Indeed, when examining this question, we found considerable segregation across the districts in the state (Stiefel *et. al.*, 2003; Stiefel *et. al.*, 2005, in press; Orfield and Lee, 2004).

To choose schools for our case studies, we began with the 2,249 elementary schools and 1,069 middle schools that had ten or more students tested in ELA and Math in 2001-02 (see Table 1). Since we were interested in only those schools that were accountable for both white and non-white students in both school years 2000-01 and 2001-02, we created a balanced panel from accountable schools in 2000-01 (780 elementary and 461 middle schools) and in 2001-02 (773 elementary and 483 middle schools); the balanced panel contained 668 elementary schools and 424 middle schools. We then limited the pool of potential schools to those that had small or non-traditional test score gaps between white and non-white students in both ELA and Math. Using these criteria, 45 elementary schools and 27 middle schools remained in the sample.

To avoid studying schools that have no test score gaps because all of their students performed poorly, we next eliminated schools where neither white nor non-white students performed at least as well as the state average. This criterion eliminated six elementary schools and six middle schools. A number of potential case study schools remained that -- although meeting the absolute minimum of 10 white and non-white students tested -- had either white or non-white students as only a very small percent of all students tested, straining the concept of an "integrated" school environment. Therefore, we instituted a relative minimum of 10 percent of students tested were white or non-white. This criterion eliminated two elementary and six middle schools.[10] In addition, we were most interested in poor schools, the group of schools largely targeted by policymakers and addressed most often by the literature, and schools where Asian students did not dominate that non-white student body.[11] For both of these criteria, we eliminated outliers: schools where less than two percent of the student body was eligible for free- or reduced-price lunches -- which eliminated 15 elementary and two middle schools -- and schools where more than 90 percent of non-white students tested were Asian, which eliminated 16 elementary and 6 middle schools. We also eliminated magnet and other selective schools. This

[10] A detailed explanation for choosing this minimum is included in Appendix A.

[11] Because white and Asian students perform similarly on standardized tests, Symonds (2003), for example, combined white and Asian students into one group and Hispanic and African-American students into a second group to calculate test score gaps within schools. Stiefel *et. al.* (2003) showed that Asian students in New York State performed as well as, or better than, white students on 4th and 8th grade ELA and Math exams in 2000-01. For this current study, we were interested not only in schools that reduced test score gaps but also, to a certain extent, schools that were not even able to calculate a test score gap due to small numbers of students in racial sub-groups. For this reason, we prioritized analyses by race rather than by relative performance groups (i.e., high performers and low performers), unlike Symonds (2003).

criterion eliminated two middle schools. Further details of the selection process are contained in Appendix A. As a final result of applying these criteria, we identified six elementary schools and five middle schools from which we were able to recruit two elementary schools and one middle school for further study.

## Qualitative Analysis

Because of the preliminary nature and necessarily limited scale of this study, no comparison group of schools was included. To help address this limitation, many of our survey questions could be compared directly to those used in two recent survey-based studies. These two surveys were chosen in particular because they both addressed hypotheses we were interested in testing, and they both included samples that we were unable to include in our own work. The first survey was undertaken by the Bay Area School Reform Collaborative (BASRC) -- as mentioned in the literature review -- and incorporated a comparison group along with a group of California schools found to be successful at reducing test score gaps (Symonds, 2003). The second study included a nationally representative sample of teachers for an extensive survey on achievement gaps and other race-related school issues for the 50th anniversary of *Brown v. Board of Education* (Markow and Scheer, 2004). By comparing our results on similar survey questions, we hope to better understand possible reasons our case study schools were successful compared to a national sample, as well as an average for a sample of unsuccessful schools.

Also, the limited number of observations did not allow us to break survey responses into elementary versus middle schools, an important distinction as it is commonly believed that students increasingly self-segregate by race as they age.[12]

To understand the factors behind the success of schools with little or no test score gaps, we used a multiple case study methodology in three schools, two from the fourth grade sample and one from the eighth grade sample. The case studies consist of data from a principal interview (protocol shown in Appendix B) and surveys of teachers (instrument available in Appendix C). We used a purposeful sampling strategy to select the schools for the qualitative analysis, choosing schools with an integrated student body, atypical or comparable test score gaps between white and non-white students, and geographic locations representing the distribution of atypical schools from the entire pool of schools within New York State. Of the schools with atypical or comparable test score gaps between white and non-white students, over 80 percent of schools with 4th grades and nearly 75 percent of schools with 8th grades were in a large downstate city. Therefore, to choose three schools for each of the 4$^{th}$ grade and 8$^{th}$ grade samples, two were chosen from the large downstate city and one from upstate.

Interviews were conducted with the principals of the three schools in May and June 2004. Two of the schools (one of the elementary schools and the junior high school) were located in the large downstate city while the other elementary school was located in an upstate small city. All three of the principals had been at

[12] Note that the limited number of observations was not due entirely to study design, rather to the limited number of schools that were eligible for inclusion in the study. Appendix A lists the criteria used to select school for further study and shows the number of schools eliminated from consideration due to the implementation of each criterion.

the school for at least three years, thus covering the time period represented by our quantitative data. Two of the principals came to their positions from administration, rather than teaching. One had worked as a central office administrator, one was a special education supervisor, while the third was a teacher and assistant principal prior to becoming principal.

Teacher surveys were distributed in May and June. All 4th or 8th grade teachers received surveys so we could capture responses of those personnel who ostensibly have the greatest impact on 4th and 8th grade test scores. The response rate for the surveys was 88.5 percent (n=23). Of the surveys returned, 91 percent came from teachers in the large downstate city. Seventy percent of the teachers responding to our survey had taught in the case study school for at least three years, covering the time period of our quantitative data. In addition, of all respondents, 30 percent had a degree in special education.

## IV. Findings from Case Studies

### School and Teacher Culture and Characteristics

**School and Teacher Characteristics.** The schools included in this study are, by definition, outliers in the population of schools in New York State. Therefore student, school, and teacher characteristics are not average on any measure, as shown in Table 2.

Compared to teachers in New York State as a whole, teachers in our sample schools had slightly more teaching experience and were more likely to be permanently certified. Compared to the other schools in the districts containing our case study schools, teachers in sample schools had a great deal more experience and were more often permanently certified.

Expenditures per pupil and median teacher salaries -- both district averages -- were lower in the sample schools' districts than in the state. Note that expenditure and teacher salary data are not available at the school level in many districts. The large downstate district, however, is an exception, thus allowing a comparison between that district's schools in our sample and the average for that district, as well as the district average across the state. The elementary school in our sample from the large downstate district had an average (mean) full-time teacher salary of $66,521; the middle school in our sample from that district had an average teacher salary of $80,621.[13]

There were no clear results from our interviews and surveys concerning class size or ability grouping and tracking. A number of teachers responded that small class sizes would help a great deal to reducing test score gaps; one teacher wrote that large classes were an obstacle to closing test score gaps, noting that "you need to have small classes and one-on-one with student/teacher relationships." Another wrote, "Classroom size is a major obstacle when trying to teach each child the necessary skills for the exam." In addition, two of

[13] Note that direct comparisons are problematic as the school-level data reported here are means and the NYS district-level data are medians, although both sources report combined salary and fringe figures for full-time teachers.

the three schools in the study were significantly overcrowded according to their principals.[14] If the successful schools we studied were indeed overcrowded, this suggests that one of the possible mechanisms making small class sizes desirable -- lack of overcrowded conditions -- was not present in at least two of our three successful schools. Note that despite this disadvantage, these schools have eliminated test score gaps.

Survey and interview results about the effects of ability grouping and tracking were inconclusive. According to teacher responses as well as discussions with school principals, only one of the two schools in the downstate large city district groups students by ability, while the school in the upstate small city district does not. Again, despite these differing classroom strategies, all three schools succeed at reducing test score gaps.

Without observing classroom conditions and without better grade-level or classroom-level data, it is difficult to assess whether individual classes themselves were overcrowded, in particular in the 4th and 8th grades. Therefore, our results are inconclusive, reflecting the state of the research on the effects of class size.

**Integration and Student Relationships.** The majority of teachers in all three schools believed that students at their school interact across race in a variety of activities in and out of school, and saw this as important for their school's success (Table 3). Most also felt that this interaction is encouraged by teachers and other staff. One elementary principal believed that the small amount of segregation that occurs is due more to language barriers than to racial tensions. The other elementary principal described student relationships by saying, "There's no differences. They don't see any differences. They're friends." This attitude was reflected not only by the other elementary principal, but also by the middle school principal who worked with the older students supposedly more likely to self-segregate.

The preponderance of teachers responding to the survey agreed that students of different backgrounds interacted with each other in and out of school. Teachers also felt that the faculty and staff at their schools encouraged this interaction:

- Nearly half the teachers surveyed (47.6 percent) "strongly agreed" that teachers in their school "encourage students of different backgrounds to work together on projects" (see Table 3).[15]
- Almost all teachers answered either "agree" or "strongly agree" (94 percent) that "students from different racial backgrounds frequently hang out together outside of class time."
- All three principals stated that they found no need to encourage students to interact; students interacted naturally on their own.

[14] The third principal specifically mentioned having excess capacity, resulting in an influx of special education students who are reducing average test scores for the school -- an issue addressed in greater detail later in this report.

[15] This and other questions were reproduced from *Education Week*'s nationally representative survey of teachers (conducted in February 2004) for the anniversary of *Brown v Board of Education*, <www.edweek.com/ew/ewstory.cfm?slug=36brownpoll.h23> and <www.edweek.com/ew/vol-23/brown_survey_final_report.rtf>; in the EdWeek survey responses, analyzed by Markow and Scheer (2004) for Harris Interactive Inc. In Markow and Scheer's results, 65 percent of teachers "strongly agreed," with 9 percent finding the question "not applicable" (Markow and Scheer added a Not Applicable option because: "If your school does not have student of different racial backgrounds, please select 'Not Applicable.'") This strong showing of support is interesting considering how many schools offer little opportunity for mixing.

Most teachers and principals supported using the diversity in their school as a way to help students learn (Table 4). The vast majority of teachers did not find "color-blind" classrooms to be as important to helping students learn as celebrating a mosaic of backgrounds. The teachers seemed to find it important to celebrate their students' differences to helping all students succeed, making the diversity of the case study schools' students a resource not available in many other schools. This point -- as well as the academic importance of diversity in student backgrounds -- was emphasized by one principal:

> Since every class has got at least five ESL kids in it, and on top of that, maybe a couple of refugees who for whatever reason haven't graduated from the ESL program. And then, on top of that, they're going to have a mix of American-born kids of different racial backgrounds…when you do the Social Studies lessons on different cultures or countries, you don't have to worry about learning about cultures you don't know because you've got so many in the room that the kids can bring their culture in and be proud of it.

Teachers believed that students mix quite a bit. They answered "sometimes" or "often" more than 90 percent of the time when asked how often students from different backgrounds interacted in six possible settings: during classes, in-between class periods, in the lunchroom, at the playground, on sports teams, and at clubs or after-school activities.[16] In addition, 100 percent of teachers answered either "somewhat" or "very important" for students to mix in four activities listed: attending school, attending classes, socializing, and participating in extracurricular activities.[17]

The vast majority of teachers in both our survey and Markow and Scheer's (2004) -- as well as the principals we interviewed -- agree that it is important for students of different backgrounds to interact. The main difference between the teachers surveyed for each of our studies was their opportunity to observe relationships between diverse students; a nationally representative sample of teachers does not have the same exposure to the diverse student body in our three study schools. We purposefully chose diverse schools that, by definition, provide more than average opportunity for teachers to observe the effects of interaction between students of different backgrounds. Perhaps this explains the large difference between the share of our study's and Markow and Scheer's teachers who agree that students of different backgrounds interact frequently. Interestingly, however, there is very little difference in the two study's results concerning how much teachers and administrators encourage this interaction.

**Teacher Expectations and Beliefs.** Teachers identified limited parental involvement and challenging family conditions, as well as a lack of student effort and lower natural ability, as primary reasons for test score gaps. They did not perceive injustice or discrimination in society, or low expectations by teachers, as explaining the gap. Table 5 shows teacher responses when asked how much they felt a specific list of factors contributes to test score gaps.

- One hundred percent of teachers said that lack of family support, challenging family conditions, and lack of student motivation explained "some" or "a lot" of the test score gap. The highest percentage

[16] In Markow and Scheer's (2004) results for *Education Week*, over 90 percent answered "often" or "sometimes" only to interaction "during classes" and "in the lunchroom."

[17] In comparison to Markow and Scheer's (2004) survey results, over 90 percent of responding teachers agreed it was important to attend school, attend classes, socialize, and participate in after-school activities with students of diverse backgrounds.

of teachers -- 91.3 percent -- responded "a lot" for "lack of student motivation/effort;" 100 percent answered either "some" or "a lot" for this option.

- Notably, 86 percent answered either "some" or "a lot" for the option, "lower natural ability."[18] Although academic background in special education might have affected teacher responses to this question, we found that while 86 percent of teachers with an academic background in special education answered "some" or "a lot," 87 percent of teachers with no such background responded "some" or "a lot."
- Less than one-third of the teachers attributed some or all of the gap to injustice or discrimination in society, while 55 percent thought that unequal access to quality education played at least some role in the test score gap.

When asked what factors would work best to reduce test score gaps (Table 6), nearly 80 percent of teachers thought that increased parental involvement would work best (see additional discussion below under the Parent and Expectations section), and 65 percent said that "increased student effort/motivation" would work best. Less than 50 percent of the teachers agreed that giving disadvantaged students equal academic opportunities would reduce the gap. No teachers agreed with the statement, "there is no solution for the test score gap" or the statement concerning the efficacy of "making schools accountable for test scores."[19] The latter result may reflect teachers' reactions against No Child Left Behind accountability measures. Some teachers in open-ended questioning explicitly mentioned that accountability measures were an obstacle to reducing test score gaps, in one case dismissing them as "political nonsense." In addition, a number of teachers and all three principals expressed concern that holding schools accountable for the performance of all sub-groups -- in particular, special education students and recent immigrants with English as a second language -- would not help students and would make otherwise successful schools look ineffective. In response to the open-ended question concerning obstacles to closing test score gaps, one teacher wrote, "Special education students are counted in the scores for schools. They should be counted separately so the school scores are not lowered." One principal noted:

> I resent, basically, the fact that they're making children who have come to this country under one year, or under two years, to take some of the state tests. I resent it because I think it's not fair. Once the child has been here two years, yes, then I can say test them, but when you're asking children that are coming from China or Korea who do not have a phonetic language and do not have even a written

[18] In Markow and Scheer's (2004) results, the highest percentage of teachers responding either "a great deal" or "some" was for "lack of family support or involvement" (97 percent) and "challenging family conditions" (96 percent). The next highest percentage of teachers responding "a great deal" or "some" (92 percent) was for "lack of motivation," with the next highest percentage dropping to 58 percent for "low expectations on the part of teachers." The highest percent answering "not at all" was 30 percent for "lower natural ability." Notably, "injustice or discrimination in society" received 44 percent of "some" or "a great deal" responses.

[19] When asked in the Markow and Scheer (2004) survey which of the three following factors would work best to reduce test score gaps, teachers most often responded affirmatively, as with our survey, (85 percent) for "increased parental involvement." The second most frequently cited option (63 percent), as in our survey, was "increased student effort/motivation," and third most cited (36 percent) was "solving economic issues outside schools," which differed from the third most cited option in our survey, ensuring "disadvantaged students have equal academic opportunities." The least frequently cited (1 percent) option was "There is no solution for the test score gap."

> language, to answer, in writing, science and social studies questions, I resent that, because it's very difficult.

Another principal added:

> I have a small school [and] … I have a lot of room in the building. So since we have room, they just give us more and more special ed. Now, it's a bit of a problem because, and I'm a special ed person in my background, that's how I came here, but the problem is… my number of general ed children is limited, and of course,… every child counts in my small numbers. So if you look at the data, it could be one or two children, and that could be a 10% change… That makes a very big difference in the way you score, but in terms of everything else, special and general ed children can do everything together…And if I'm not mistaken, I don't have it in front of me, but our total school number of successful children that score level three or four I think it's around 49% but if you tease out our special education population, it's like 73%.

Teachers' apparent lack of belief in the utility of accountability standards for reducing test score gaps could be based on the unanimous agreement among teachers that lack of family support, challenging family conditions, and lack of student motivation explained "some" or "a lot" of the test score gap. Teachers might perceive that accountability measures do not address what they feel are the most important obstacles to reducing test score gaps. Accountability standards are arguably the option most threatening to teachers, and thus their responses might not be surprising. However, dismissing these concerns is not the design of this study.

**Diverse faculty and staff.** Although the case study schools' surveyed teachers were 60 percent female, and 91.3 percent white (with the remaining 8.7 percent Asian), teachers tended to agree that people of color were visible in the staff of their schools, that closing test score gaps was a priority for the school (in particular, by emphasizing improving the performance of all students), and that teachers worked in an environment of collaboration. When asked the extent to which they agreed with the statement, "People of color hold positions of power in this school," the majority (63.1 percent) "strongly agreed" or "agreed."[20]

Teachers' responses in both our survey and Markow and Scheer's (2004) emphasize the widespread belief of teachers that home factors matter a great deal in determining student performance. But a much higher percent of the nationally representative sample of teachers in Markow and Scheer's study believed lower teacher expectations explained test score gaps than did the teachers in the successful schools in our study. Diversity of faculty and administrative staff also might play a role in reducing test score gaps, as teachers in our successful schools as well as those in BASRC's successful schools study (Symonds, 2003) agreed that people of color held positions of power in the school.

[20] This question was reproduced from the Bay Area School Reform Collaborative's (BASRC) achievement gaps teacher survey for their 2003 study "After the Test: How Schools Are Using Data to Close the Achievement Gap" by Kiley Walsh Symonds <www.basrc.org>. When asked this question in the BASRC survey, teachers in gap-closing schools (i.e., schools that have reduced test score gaps among students of different racial/ethnic backgrounds) "strongly agreed" or "somewhat agreed" 71 percent compared to 37 percent of teachers in non-gap-closing schools.

**Leadership.** The principals we interviewed are all dynamic instructional leaders who firmly believe in supporting their teachers. All three principals describe themselves as having an open door policy, and giving teachers the flexibility to try new approaches. One principal said:

> People know they can reach me here, or at home, if they have a suggestion. I've learned of the possibility, if it's good for the children, it's good for the school. You have to be a risk taker. You've got to learn to try things, to see what works for the kids. If the teacher has a good idea, and she's willing to go for it, I'll support them if it makes sense because, let them try it. If it doesn't work, go to plan B.

**Principal Expectations and Focus.** Teachers also stressed the importance of their principal's leadership. The majority of teachers surveyed believe their principal values their expertise (see Table 7). They also agreed that the principal knows and cares for children, and felt their principals were effective leaders who made the school run smoothly. These are all characteristics of strong leaders.

Despite these positive characteristics of the leadership of their schools, teachers also felt they had little say in the way their school is run. Most teachers did not believe they have much involvement in deciding school policy. Thirty-two percent of teachers reported having at least some involvement in planning how discretionary school funds should be used, while almost 26 percent said they had at least some involvement in establishing the curriculum and instructional program or determining the content of in-service programs. Only 4 percent said they had some involvement in hiring new professional staff. Perhaps the cost of having strong leaders is the loss of teacher input into school policy and hiring decisions.

**Staffing, Goals, and Standards.** Two of the principals gave credit to their predecessors for laying the foundation for their current success. When asked what the school was like before she came, one principal said: "It was a good school. The principal was certainly a man who worked very hard and did a good job, and he laid a foundation for what we're doing now." Regardless of the state of the school when they came in, however, all three principals highlighted the importance of being able to bring in their own staff. At all three schools, the current principal has hired over half the staff. One principal said:

> The ones that are here are here because they, you know, working with this kind of kid works for them. It fulfills their needs. I've hired about 80% of the folks that are here. Usually we'll end up getting teachers who somehow have gotten experience in our buildings and who we know like us and we like them. Long-term subs, frequent subs, that kind of thing. But supply and demand is such here that we have the luxury of hiring people that we have experience with.

When teachers were asked to respond to the statement, "Closing the test score gap between white students and students of color has been a primary goal for school leaders," 55 percent either "agreed" or "strongly agreed." Teachers were also asked to what extent "school leaders set measurable goals for closing the test score gap," and, again, the majority agreed, with 58 percent either "strongly agreeing" or "agreeing." However, only a little over half the teachers either "strongly agreed" or "agreed" that "the school community developed and agreed upon an explicit definition of 'equity.'"

To summarize the role of leadership in reducing test score gaps, teachers acknowledged having a great deal of confidence in their principals, perhaps because most teachers we surveyed also were hired by the principal we interviewed. Principals were given flexibility to hire the teachers they thought best met the needs of their students, and in return, teachers seemed to believe in the ability of the principals to lead the school to success, despite their feeling they are not involved in school policy decisions. Strong leaders who have a stake in the hiring of their staff as well as a focus on reducing test score gaps seem to be more successful.

## Use of Data in Schools

**Identifying Low-Performing Students.** As other studies have found, teachers in our sample school regularly use data to understand and identify weak student skill areas. Almost 50 percent of the teachers reported using data to understand the skill gaps of low-achieving students at least a few times per month. The majority of teachers reported that they administer ongoing assessments throughout each month to better identify low-performing students. In addition, some principals noted that they target special services to students identified as needing extra attention:

> We gave every special ed class we thought needed it testing preparation that would ensure their ability to be able to do the test. In other words, I don't think a child should be penalized in math, social studies or science because they can't read. If that child could have a test read to them, then we try to see that we're able to get that modification in for them.

**Acting on Data.** Discussions of the use of data highlight a philosophical issue: once data provide insight into students who might need extra help, is it more important to provide equal attention to all students, or to provide more help to a few students even if other students receive less attention? Table 8 shows that the majority of teachers feel the latter is most important; not a surprising result considering the federal, state, and district focus on using data to identify low-performing students.

While principals have varying levels of ability in understanding data, they all say they use some type of computer program to help them and their teachers use the findings from assessments (including their own ongoing assessments). The principal's knowledge about data plays a big part in how the data are used within the school and how data are translated to teachers. One principal said:

> I think we probably use data more than most schools. That's been a big push in our district. And we've bought into it. I mean, top-down innovations are no good if people in the field don't buy into it. And I think, my ability with data -- now I used to be the district computer service director for 11 years -- I mean I understand data well enough that I can take what the district sends us and help the teachers make some sense of it. What the district sends us is good, but it's not really action-able or available. It's a stack of reports with a bunch of numbers. That's scary to your typical teacher.

Although each school uses a different system, they all have several characteristics in common. One principal describes the system by saying:

> It's an integrated learning system…It's one of these things where you put the kid's name in, what grade level they're at in reading and math, it gives them tutorials and drills, and every time the kid responds, it analyzes the response and figures out what to do next. It's like a real smart system. One of the things that's kind of neat is that when the kids are working the teachers can just sort of scan the

> screen to see what the kid's on. It's like an instant assessment. And then there are reports they can print out that give more details.

One of the stated goals of these programs is to increase parental involvement in student learning by displaying performance results. Unfortunately, as all three principals pointed out, few of their school's parents have access to or knowledge of computers, making it difficult to incorporate parental involvement into ongoing assessment systems. As one principal said: "We're not talking about an average middle class home where two parents are able to help their children and parents have computers and all kinds of technology."

In addition to the state and district mandated yearly assessments, the principals say their teachers administer other assessments at least a few times a year to assess how students are progressing. Fifty-five percent of teachers said they "use data to understand the skill gaps of low-achieving students" less than once a month, with 45.4 percent saying they use data at least once a month. Of those users, 13.6 percent said they use data at least a few times per week.[21] And 100 percent of teachers reported administering ongoing assessments to students (Table 9), with almost 83 percent saying they did this either a few times per week or per month.[22]

Teacher communication plays a prominent role in identifying students in need of additional assistance. Eighty-seven percent of teachers said they talk with their colleagues at least a few times a month about low-performing students.[23] One principal explained the school's active pupil-personnel teams (PPT, which, although not unique to this district, involved components not included in other schools' teams[24]). The principal said that these meetings were attended by a variety of school personnel -- the principal, assistant principals, a guidance counselor, psychologist, and teachers (general and special education) -- helping them to identify individual students early before any specific problems occur or escalate. She noted the important role of teachers in identifying students with an educational or social/interaction issue early in the school year, and in assigning a specific person to follow up on the particular student and report back to the team regularly. She continued,

> The thing that I really like the most is at the meeting a follow up assignment is given to one specific person. Instead of just saying, okay, that child, okay, you'll look into it… So someone knows that

[21] When asked the same questions in the BASRC study, Symonds (2003) found that teachers in gap-closing schools frequently used data to understand gaps in the skills of low-achieving students. Teachers responded "A few times a week" or "a month" 64 percent in gap-closing schools but only 21 percent in non-gap-closing schools, and "A few times a year" in 36 percent of the gap-closing schools compared to 74 percent in non-gap-closing schools.

[22] Teachers in the BASRC survey (Symonds, 2003) responded to this same question "A few times a week" or "a month" 91 percent in gap-closing schools but only 79 percent in non-gap-closing schools, and "A few times a year" in nine percent of the gap-closing schools compared to 21 percent in non-gap-closing schools

[23] This question was reproduced from the BASRC teacher survey (Symonds, 2003); teachers in the BASRC survey responded "a few times a week" 32 percent of the time in gap-closing schools and five percent in non-gap-closing schools; "a few times a month" 45 percent of the time in gap-closing schools compared to 48 percent in non-gap-closing schools; and "a few times a year" 23 percent of the time in gap-closing schools compared to 47 percent in non-gap-closing schools.

[24] The principal emphasized that the weekly PPT was not intended as a road to special education, rather as a mechanism for addressing students' social and educational problems early. The team meets during a common preparatory period for the teachers involved. The meetings are used to identify emerging problems in students. Most importantly, one team member is assigned to follow up on the student's problem (social or academic) and to report back to the committee every week until the problem is resolved. If problems escalate, the principal already understands the student's background, and knows whether, for example, to avoid calling the parents because of the potential for a domestic dispute. If there is a pattern of problems, the correct services can be delivered to the student, and one team member will report weekly on the progress of the student. During our visit, a principal-in-training was observing the principal and school operations, and in particular, the pupil-personnel team about which he was writing a research paper.

> they're responsible and accountable to bring that information back, and it also personalizes it for that child so, I may not know who this child is by sight, but someone does, and someone is checking with their teachers and other people in the building to find out what's going on… And the other day we really saw the effect of it when an incident happened, and we knew everything about this kid. And I never saw the kid. I've never seen him before that day, but I knew exactly, you know, not to call mama.

This specific program helped to highlight the importance of identifying specific student needs (social or academic) as early as possible. These early interventions appear to help school leaders target limited school resources to where they are needed most, which helps individual students succeed while creating a less chaotic school environment so all other students can succeed as well. Results from our teacher survey underscore the significance of student assessment data to reducing test score gaps, through the identification of students who need extra academic assistance. But larger quantities of data alone are not the solution: the effects of data are increased if teachers assess student progress on an ongoing basis and are able to interpret and discuss the data they receive. These results are reflected not only in our study's results, but also in the results of the BASRC comparison of gap-closing and non-gap-closing schools (Symonds, 2003).

**Professional Development and Instruction.** To support the activities teachers identified as important to reducing test score gaps and improving the performance of low-performing students, teachers more frequently reported receiving professional development in literacy instruction, as well as in the analysis and linking of data to instructional strategies (see Table 10).[25] In a separate question, when we asked how often teachers received professional development about literacy instruction, 31.8 percent responded "a few times a month" and 54.6 responded "a few times a year." However, 13.6 of the teachers said they had not received any professional development in literacy instruction in the past year.[26,27]

Fifty-four percent of teachers reported receiving "professional development about analyzing data on low-performing students" less than once a month, with 45.5 percent reporting receiving such training at least once a month. Only 4.5 percent reported "never" receiving such training.[28] Fifty-nine percent of teachers reported receiving "professional development about linking data on low-performing students with effective

[25] It is not surprising, however, that over two-thirds of teachers said they received professional development in literacy instruction in the past year, given that most of the teachers who responded to the survey are in a large downstate district that implemented a new literacy curriculum in the 2003-04 school year that included increased professional development in teaching literacy skills.

[26] When asked this question in BASRC's survey, 50 percent of teachers responded "a few times a week" or "a month" in gap-closing schools and 26 percent in non-gap-closing schools, and 35 percent "a few times a year" in gap-closing schools compared to 74 percent in non-gap-closing schools.

[27] Surprisingly, only 39 percent said they received professional development about teaching math over the past year (again, see Table 10), since a new math curriculum with increased professional development was also implemented in the large downstate district in the same school year.

[28] Symonds (2003) reported teacher responses to this same question in the BASRC teacher survey; teachers responded "a few times a month" 50 percent in gap-closing schools compared to 31 percent in non-gap-closing schools; "a few times a year" in 50 percent of the gap-closing schools compared to 58 percent in non-gap-closing schools; and "Never" in zero percent of the gap-closing schools compared to 11 percent in the non-gap-closing schools.

instructional strategies" less than once a month, while 41 percent said they received such training at least once a month.[29] One principal described the professional development she provided around performance data:

> The initial results come out, I'll do it in a large group and then spend time with each grade, at the grade conference, going over particulars. To be aware of the needs of all the children, not to use it, but to define instruction.

In contrast, 82.6 percent of teachers did not receive professional development training in "multicultural diversity issues." When asked about providing professional development around multicultural diversity issues, one principal acknowledged the difficulty of teaching students of so many different backgrounds in large classes and said her teachers had recently had professional development on this topic. Another principal, however, said:

> There really is not much of a need for that. They know it. When I hire staff, I look for a new teacher, I look for someone who's going to have that under his belt, otherwise, they don't belong here.

This analysis of professional development activities shows the potential importance of training in literacy skills, as well as in the analysis and interpretation of data, to enable teachers to improve student performance and reduce test score gaps. Most teachers in our study and in the gap-closing schools of the BASRC study receive professional development concerning literacy instruction and in linking data to instructional strategies throughout the year. These skills seem to be much more important than multicultural and diversity issue training to reducing test score gaps.

**Parents and Expectations.** The majority of teachers believed that students and families, rather than social factors, are the biggest contributors to the test score gap. One teacher wrote in response to the open-ended question, "What do you think schools should do to close test score gaps?"

> More parent involvement. I am responsible for a child for three hours per week for 38 weeks. The parents are responsible 24/7 for 18 years. If a child cannot read, am I to blame?

In response to the second open-ended question, "What do you think are the biggest obstacles to schools' closing test score gaps?" another teacher wrote, "I believe where these children come from and go home to every day is a huge obstacle."

Teachers believed that although they and their colleagues all make efforts to engage parents, many parents were not involved in their children's schooling. Responses showed a disjuncture between what teachers think they do and how effective they are with parents (Table 11). While 95.5 percent of teachers said they regularly communicate with parents about how their children can learn, less than two-thirds believed that

[29] Again, this question was reproduced from the BASRC teacher survey; teachers in the BASRC survey responded "a few times a week" five percent in gap-closing schools and zero percent in non-gap-closing schools; "a few times a month" 45 percent in gap-closing schools compared to 26 percent in non-gap-closing schools; "a few times a year" in 45 percent of the gap-closing schools compared to 63 percent in non-gap-closing schools; and "never" in five percent of the gap-closing schools compared to 11 percent of the non-gap-closing schools.

parents have an understanding of what is expected and can offer suggestions on how best to help their children.

When asked about parental involvement in the school, teachers tended to believe that parents do not get involved unless they are asked directly. When teachers take the initiative, they feel more parents get involved: two-thirds of teachers felt that parents of their students attended parent-teacher conferences when requested. However, only 13.7 percent of teachers felt that even half the parents of their students requested a parent-teacher conference to talk about their child. Parental attendance at school-wide events was even less, with only 20 percent of teachers reporting that at least half the parents attended these events. Recall that when asked what factors would work best to reduce test score gaps (Table 6), almost 80 percent of teachers thought that increased parental involvement would work best. In addition, one principal said:

> The lives of many of our parents are chaotic. When you're dealing with a poor demographic, there are certain things that go hand in hand with it. Alcohol abuse, neglect, mess at home.

Another principal added:

> As I figured, there's difficulty in getting the parents to come in because of the nature of our population. We have a large immigrant population. Many parents are not even able to read and write in their own languages and they are faced with the problems of new immigrants. Some of them value education, not all. And so their children come to school not having knowledge of English, not having may experiences and …that's a very big problem we face. The parents don't come in because they're not legal and they have financial problems and family problems. There's a myriad of difficulties.

The above quote also highlights the role that immigrant students play in the schools. All three principals said that the non-white population at their school is comprised predominately of immigrant students. While these parents tend not to be very involved in school life, they do care strongly about education.

Having such a large immigrant population clearly impacts how principals target parents. One of the principals highlighted the importance of being able to find people who speak the same language as the parents. She had people on staff who could translate, and has made connections in the community that she can rely on when necessary. However, these practices were difficult to identify at the other two schools, since most of the students were not from the neighborhood.[30]

Unlike the principal who said her immigrant students are faced with chaotic family lives, another principal thought that the main difference between the refugee population and native-born children of color is that while the immigrant families may be poor, they may have more stability in their lives than many of the native-born families. He said:

> One of the good things about the refugee population is that they are two-parent families. Single moms don't get out. I don't know if you've thought of it that way before, but if you're a single mother in the Congo, you're not getting out. Or any of these other places. You need, I use the metaphor of a filter

[30] Although the three case study schools were neighborhood schools, intended to pull students from the local area, because of a large number of open spaces within the schools and private school attendance among families in the neighborhoods the schools served, these schools received a large number of students bused in from outside the neighborhood schools' service area.

> here in terms of why I get the refugees I get. There is some filtering mechanism that allows the ones that get out to get out. They have some qualities that got them out. Intelligence, initiative, perseverance, connections, you know, you name it. They had some qualities that got them out. Once they get out, those qualities serve them well when they're here. And they value education and they don't want to hear about the principal calling home and saying, hey, come and get your kid. Take a look at the test scores from my Somali families. I mean, not only do they have families, they have a tribe. They've got this guy who's sort of the de facto chief that everybody looks up to in addition to their parents. There's like a quasi-tribal structure within the Somali community here. If you think it takes a village to raise a kid, they sort of got it, a village-like situation going for them. So if somebody's mother can't help the kid with something, somebody else's aunt can. If you were to cull out my Somali scores you'd find that they out-performed the white kids.

While most teachers agreed that reducing the test score gap is a priority for both their school and their district, they were divided about the strength of the school's response to reducing the test score gap (see Table 12). Only half of the teachers agreed or strongly agreed that the school has made explicit the importance of reducing the test score gap in their school or finding ways to achieve this goal. Teachers were asked to what extent "school leaders set measurable goals for closing the test score gap," and, again, the majority agreed, with almost 16 percent "strongly agreeing" and 42.1 percent "agreeing."[31,32,33]

From teacher and principal responses, there seemed to be a number of parent, peer, and neighborhood factors that contribute to student performance and test score gaps. The respondents in our study defined these factors as largely outside the control of the school. Although school leaders and teachers thought that parental involvement was important to improving student performance, many found that parents were unable to be involved in their children's schools. However, direct parent involvement in schools is not the only kind of involvement parents can undertake in their children's education. Parents, as well as teachers and principals, can make clear their expectations for all students to achieve at high levels. Principal surveys, in particular, showed that this type of communication of parental expectations occurred most often in recent immigrant households. Unfortunately, there are few policy recommendations stemming from this finding that can be provided to schools without populations of recent immigrants; reductions in test score gaps due to the presence of immigrant children are not reproducible in districts with little immigration. What may be reproducible is the direct communication of teacher and principal expectations that all students achieve at high levels, especially in poor districts where parents are not able to get to a school for in-person visits (as was the case with our case study schools). Note again the quote on page 41, where one principal illustrated the differences he saw in native-born and recent immigrant families, in particular the Somali families with

[31] In the BASRC survey, when asked this question, 45 percent of teachers "strongly agreed" in gap-closing schools compared to 11 percent in non-gap-closing schools; 50 percent "somewhat agreed" in gap-closing schools compared to 63 percent in non-gap-closing schools.

[32] When asked to what extent they agreed or disagreed with the statement, "School leaders encourage or lead school-wide inquiry into the test score gap," 55 percent either "agreed" or "strongly agreed." When asked this question in the BASRC survey, 100 percent of teachers in gap-closing schools either "strongly agreed" or "agreed" compared to 79 percent of teachers in non-gap-closing schools.

[33] Teachers were asked to what extent "school leaders provide structured opportunities for faculty to discuss race and ethnicity." Almost 17 percent "strongly agreed" and 16.7 percent "agreed." In the BASRC survey, when asked this question, 14 percent of teachers "strongly agreed" in gap-closing schools compared to 21 percent in non-gap-closing schools; 77 percent "somewhat agreed" in gap-closing schools compared to 16 percent in non-gap-closing schools.

successful students in his school. This principal's comment points out a major difficulty in examining test score gaps. What role do recent immigrants play in reported white/non-white test score gaps? There has been some media attention about this phenomenon, but little academic research, due largely to data constraints. Some researchers have made efforts to examine the characteristics of immigrant students in New York City, where there are data available as well as enough immigrant students to allow for reliable analyses of the data (Ellen *et. al.*, 2002; Conger, 2003, unpublished; Stiefel *et. al.*, 2002).

Placing a clear priority on reducing the test score gap also seems to help focus the efforts of the entire school community. In both the BASRC study and our own, the majority of teachers agreed that both the school and district leadership were placing a priority on reducing test score gaps, and putting in place measurable goals for reducing gaps.

## V. Conclusions

Teachers' and principals' belief that home factors have a profound effect on student performance could imply that school faculty and staff are abdicating responsibility for improving student performance. However, what seemed to matter was how teachers and principals acted on this belief. The school faculty in our study appeared to take on the challenge of home factors by improving their own instructional practices, using data and on-going assessments to better identify students who need the most help and attention, and taking advantage of strong leadership to focus instruction, avoiding unproven fad practices, and allowing teachers to be creative while still adhering to state standards.

Although most teachers expressed an interest in -- and belief in the importance of -- more parental involvement, principals pointed out that, despite parents' inability to be more involved in school (because of a lack of transportation options, unusual work hours, etc.), parents were involved in other ways. All three principals noted the importance of parents' expectations for their children. Principals perceived that parents made clear to their children that they were expected to do well in school, and that parents would not appreciate being called to speak with the child's teachers or principal because of poor performance. Students seemed to be reacting positively to these expectations.

Teachers and principals in our study believed that students benefited from relationships with students of different backgrounds. Unfortunately, most schools in New York State do not offer this opportunity because of segregation between white and non-white populations. Leaders in the schools we studied stated that they made little effort to get students to interact -- students did this "on their own," implying that the level of interaction needed to bring about good student outcomes is outside the control of schools.

The successful schools we studied all had a strong immigrant presence. However, a strong immigrant presence does not make our successful schools different than most other (less successful) schools in the districts we studied; both school districts containing our case study schools have large immigrant populations. Rather, the key component of immigration status may be parental expectations of student performance that could have important peer effects within, if not across, schools.

Teachers and principals also highlighted the importance of data, on-going assessments, and data analysis tools to help them identify low-performing students. Teachers may find this the best way to overcome home factors that may place some students at an academic disadvantage. Also, teachers' use of on-going assessments improved the impact of data in identifying not only the students who need extra attention, but also the areas where students need help (especially those students who need help in only one of many fields in which they are tested.) Principals in this study allowed teachers to be creative in following up on what they learned from the data. This belief in the abilities of teachers to address the needs of their students may stem from the principals' stated ability to hire teachers they thought best met the needs of the students in that particular school. Teachers also held high expectations for all their students and communicated those expectations regularly.

Despite the economic disadvantage of most of the students in the schools we studied, a number of school strategies appear to help overcome the negative effects of home factors. These potentially beneficial characteristics of schools and personnel include: school leaders who believe in their teachers' abilities enough to let them be creative in meeting their students' needs; data and analysis systems that help teachers identify which students need extra help and in what areas; high expectations for student performance on the part of parents, teachers, and principals; and the natural interaction between students of different backgrounds that may contribute to building an environment conducive to student success. Although further research would strengthen these conclusions, the additional evidence provided by these successful schools in New York State could be useful to formulate future gap-closing policy across the state.

## Directions for Future Research

Based on the findings from this study, a number of directions for future research are suggested:

**Students.** Students may benefit from interacting with students of different backgrounds. To better understand the importance of student interaction and the mechanisms involved in improving student performance, direct observation of student activities in a variety of settings would be beneficial. Direct observation would allow researchers to verify the perceptions of teachers and school leaders, but also help to disentangle the mechanisms involved in interaction's role in improving student performance. Do high levels of diverse student interaction change peer expectations for performance? Does interaction improve study habits for all students?

Students also may react in a positive manner to parent, teacher, and principal expectations. To further explore this hypothesis, a student survey concerning their perceptions of teacher, principal, peer, and parent expectations would be useful. Additionally, the differing expectations of children of recent immigrant versus non-immigrant families also should be explored further.

**Parents.** Similarly, survey work exploring the expectations of parents -- in particular, immigrant versus non-immigrant parents -- could highlight differences in the perceived expectations of parents, students, and teachers.

**Teachers and Principals.** Because teachers and principals highlighted the effects of home factors on student performance, additional questioning in this area could be beneficial. How do teachers think they can overcome potential negative influences on home factors? How do they feel about schools as providers of social services beyond education? Teachers and principals highlighted the importance of data, on-going assessments, and data analysis tools to help them identify low-performing students. Teachers may find this to be the best way to overcome home factors that may place some students at an academic disadvantage. Further survey work with teachers could test this hypothesis.

Also, direct classroom observation could validate teacher responses concerning instructional practices, and provide further insight into student interactions within classroom settings. Adding more schools to the sample also would yield more reliable conclusions. The addition of a comparison group of schools that were not successful in reducing test score gaps would also strengthen the findings of current and future research. A comparison group would enable us to see whether practices and beliefs are indeed different in successful schools, rather than random effects or practices undertaken in all schools. This current effort -- as well as future work strengthening conclusions drawn from this current study -- could show that there are replicable practices schools can employ to change outcomes for their students.

## VI. References

Berliner, David and Bruce Biddle (1995). *The Manufactured Crisis: Myths, Fraud, and the Attack on America's Public Schools* (Reading, MA: Addison-Wesley).

*Brown, Oliver et. al.. v. Board of Education of Topeka* (347 U.S. 483).

Burtless, Gary (1996). *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success* (Washington D.C.: the Brookings Institution).

Calhoun, Dan, Scott Linner, and Glen Szymoniak (2004). "Using Data to Improve Student Performance." Presentation at the Alaska NCLB Winter Conference, January 7, 2004 < http://www.educ.state.ak.us/nclb/Presentations/Using%20Data%20to%20Improve%20Student%20Performance.ppt> (July 19, 2004).

Card, David and Alan Krueger (1996). "School Resources and Student Outcomes: An Overview of the Literature and New Evidence from North and South Carolina," *Journal of Economic Perspectives,* v. 10, n. 4: pp.31-50.

Chubb, John and Terry Moe (1990). *Politics, Makers, and America's Schools* (Washington, DC: Brookings Institution).

Coleman, James, Ernest Campbell, Carol Hobson, James McPartland, Alexander Mood, Frederick Weinfeld, and Robert York (1966). *Equality of Educational Opportunity* (Washington, DC: US Government Printing Office).

Colorado Department of Education (2003). "Closing the Achievement Gap Coalition," <www.cde.state.co.us/cdecare/clg_home.htm> (July 20, 2004).

Conger, Dylan (2003, unpublished). "Ethnic Segregation Within Schools: Prevalence and Possible Sources in New York City." Presented at the New York University's Wagner Graduate School of Public Service Doctoral Colloquium, December.

Connecticut State Board of Education (2003). "Closing the Achievement Gaps: Removing the Barriers to Preschool in Connecticut," released November 2003, <www.state.ct.us/sde/deps/Early/PreschoolReport.pdf> (July 20, 2004).

Cook, Michael and William Evans. (2000). "Families or Schools? Explaining the Convergence in White and Black Academic Performance." *Journal of Labor Economics,* v. 18, n. 4: 729-754.

Cox, Kathy (2003). "672 Georgia Title 1 Schools Meet Accountability Goals; Cox Details Plans to Improve High Priority Schools and Eliminate Achievement Gap," press release, August 5, 2003, Georgia State Superintendent of Schools, <www.doe.k12.ga.us/doe/media/03/080503.asp> (July 20, 2004).

Ellen, Ingrid (1999). "Spatial Stratification Within US Metropolitan Areas," in Alan Altshuler, ed., *Governance and Opportunity in Metropolitan America* (Washington, DC; National Research Council): 192-212.

Ellen, Ingrid (2000). *Sharing America's Neighborhoods: The Prospects for Stable Racial Integration* (Cambridge, MA: Harvard University Press).

Ellen, Ingrid Gould, Katherine O'Regan, Amy Ellen Schwartz, and Leanna Stiefel (2002). "Immigrant Children and New York City Schools: Segregation and Its Consequences." *Brookings-Wharton Paper on Urban Affairs 2002*: 183-214.

Ferguson, Ronald (1998a). "Teachers' Perceptions and Expectations and the Black-White Test Score Gap" in Christopher Jencks and Meredith Phillips, eds., *The Black-White Test Score Gap* (Washington DC: Brookings Institution): 273-317.

Ferguson, Ronald (1998b). "Can Schools Narrow the Test Score Gap?" in Christopher Jencks and Meredith Phillips, eds., *The Black-White Test Score Gap* (Washington DC: Brookings Institution): 318-374.

Freeman, Richard and Harry Holzer (1986). *The Black Youth Employment Crisis,* eds. Richard Freeman and Harry Holzer (Chicago: University of Chicago Press).

Fryer, Roland and Steven Levitt (2004). "Understanding the Black-White Test Score Gap in the First Two Years of School," *Review of Economics and Statistics,* v. 86, n. 2 (May): 447-464.

Gilyard, Keith (1991). *Voices of the Self: A Study of Language Competence* (Detroit: Wayne State University Press).

Greenwald, Rob, Larry Hedges, and Richard Laine (1996). "The Effect of School Resources on Student Achievement." *Review of Educational Research*, v. 66, n. 3 (Fall): 361-396.

Hanushek, Eric (1971). "Teacher Characteristics and Gains in Students Achievement -- Estimation Using Micro Data." *American Economic Review*, v.61, n.2: 280-288.

Hanushek, Eric (1981). "Throwing Money at Schools." *Journal of Policy Analysis and Management*, v.1, n.1 (Fall): 19-31.

Hanushek, Eric (1986). "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature,* v. 24, n. 3: 1141-1177.

Hanushek, Eric (1992). "The Trade-Off Between Child Quantity and Quality," *The Journal of Political Economy*, v.100, n.1 (February): 84-117.

Hanushek, Eric (1997). "Assessing the Effects of School Resources on Student Performance: An Update." *Educational Evaluation and Policy Analysis*, v. 19, n. 2 (Summer): 141-164.

Hanushek, Eric (2002). "Publicly-Provided Education." Working Paper 8799, National Bureau of Economic Research. <http://dsl.nber.org/papers/w8799.pdf> (February 25, 2003).

Hanushek, Eric, John Kain, and Steven Rivkin (2001). "New Evidence About Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement." National Bureau of Economic Research Working Paper 8741, December <http://papers.nber.org/papers/w8741.pdf> (July 19, 2004).

Hedges, Larry, Richard Laine, and Rob Greenwald (1994). "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes." *Educational Researcher,* v. 23, n. 3 (April): 5-14.

Hedges, Larry and Amy Nowell (1998). "Black-White Test Score Convergence since 1965," in Christopher Jencks and Meredith Phillips, eds., *The Black-White Test Score Gap* (Washington DC: Brookings Institution): 149-181.

Hedges, Larry and Amy Nowell (1999). "Changes in the Black-White Gap in Achievement Test Scores," *Sociology of Education*, v. 72, n. 2(April): 111-135.

Herrnstein, Richard and Charles Murray (1994). *The Bell Curve: Intelligence and Class Structure in American Life,* (New York: Free Press).

Horne, Tom (2004). "Arizona State Education Superintendent's First Annual State of Education Speech," <www.ade.az.gov/administration/superintendent/articles/2004StateofEducation.doc> (July 19, 2004).

Huggins, Elise and Mary Beth Celio (2002). "Closing the Achievement Gap in Washington State: Holding Schools Accountable for Equity." Center on Reinventing Public Education at the University of Washington, November, <www.k12.wa.us/accountability/reports/FINALAchievemGapReport.pdf> (February 10, 2004).

Krueger, Alan (1999). "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics,* v. 114, n. 2 (May): 497-532.

Jencks, Christopher and Meredith Phillips, eds. (1998). *The Black-White Test Score Gap* (Washington DC: Brookings Institution).

Jepsen, Christopher and Steve Rivkin (2002). "What is the trade-off between smaller classes and quality?" National Bureau of Economic Research Working Paper, September <http://papers.nber.org/papers/w9205.pdf> (July 19, 2004).

Lareau, Annette and Erin McNamara Horvat (1999). "Moments of Social Inclusion and Exclusion: Race, Class, and Cultural Capital in Family-School Relationships." *Sociology of Education*, v. 72, n. 1 (January): 37-53.

Markow, Dana and Marc Scheer (2004). "Brown v. Board of Education, Fifty Years Later: A Survey of Teachers and Students," *Education Week*, April, <www.edweek.com/ew/vol-23/brown_survey_final_report.rtf> (May 14, 2004).

McGee, Glenn (2001). "Closing Illinois' Achievement Gap: Lessons from the 'Golden Spike' High Poverty High Performing Schools," <www.isbe.net/board/meetings/june02meeting/closinggap.pdf> (July 20, 2004).

Murnane, Richard (1983). "Understanding the Sources of Teaching Competence: Choices, Skills, and the Limits of Training." *Teachers College Record*, v.84, n.3 (Spring): 564-569.

National Center for Education Statistics (2003). "Overview and Inventory of State Education Reforms: 1990 to 2000," <http://nces.ed.gov/pubs2003/2003020.pdf> (July 20, 2004).

NEA Foundation for the Improvement of Education (2003). "Using Data About Classroom Practice to Improve Professional Development for Educators." <www.nfie.org/publications/usingdataIB.pdf> (July 20, 2004).

New York State Department of Education (2002). *New York State School Report Card.* <www.emsc.nysed.gov/repcrd2002/database.html> (July 8, 2002).

Oakes, Jeannie (1985). *Keeping Track: How Schools Structure Inequality* (New Haven, CT: Yale University Press).

O'Connell, Jack (2003). "Prepared Testimony of State Superintendent Jack O'Connell to the Senate Judiciary Committee in Opposition of Proposition 54, August 27, 2003," California Department of Education Press Release <www.cde.ca.gov/nr/ne/yr03/yr03rel91.asp> (July 20, 2004).

Orfield, Gary and Susan Eaton (1996). *Dismantling Desegregation: The Quiet Reversal of Brown v. Board of Education,* New York: The New Press.

Orfield, Gary (2001). "Schools More Separate: Consequences of a Decade of Resegregation." <www.civilrightsproject.harvard.edu/research/deseg/Schools_More_Separate.pdf> (July 18, 2003).

Orfield, Gary and Chungmei Lee (2004). *Brown at 50: King's Dream or Plessy's Nightmare?* <www.civilrightsproject.harvard.edu/research.reseg04/resegregation04.php> (March 3, 2004).

Perry, Theresa, Claude Steele, and Asa Hilliard, III (2003). *Young, Gifted, and Black: Promoting High Achievement Among African-American Students* (Boston: Beacon Press).

Phillips, Meredith, James Crouse, and John Ralph (1998). "Does the Black-White Test Score Gap Widen After Children Enter School?" in Christopher Jencks and Meredith Phillips, eds., *The Black-White Test Score Gap* (Washington DC: Brookings Institution): 229-272.

Rivkin, Steven (2000). "School desegregation, academic attainment, and earnings." *Journal of Human Resources,* v. 35, n. 2 (Spring): 333-346.

Rivkin, Steven, Eric Hanushek, and John Kain (1998). "Teachers, schools, and academic achievement." National Bureau of Economic Research Working Paper 6691, August <http://papers.nber.org/papers/w6691.pdf> (July 19, 2004).

Rothman, Robert (2001-02). "Closing the Achievement Gap: How Schools Are Making It Happen," *Challenge Journal: The Journal of the Annenberg Challenge*, v. 5, n. 2 (Winter), <www.annenberginstitute.org/Challenge/pubs/cj/cjv5n2.pdf> (July 19, 2004).

Rothstein, Richard (2004). *Class and Schools: Using Social, Economic, and Educational Reform to Close the Black-White Achievement Gap* (Washington, DC: Economic Policy Institute).

Schiller, Bradley (1989). *The Economics of Poverty and Discrimination*, 5th edition (Englewood Cliffs, NJ: Prentice Hall).

Siddle Walker, Vanessa (2001). "African American Teaching in the South: 1940-1960." *American Educational Research Journal,* v. 38, n.4: 751-779

Snipes, Jason, Fred Doolittle, and Corinne Herlihy (2002). "Foundations for Success: Case Studies of How Urban School Systems Improve Student Achievement," Manpower Demonstration Research Center and the Council of the Great City Schools, <www.mdrc.org/publications/47/full.pdf> (July 19, 2004).

State of Florida/US Department of Education (2003). "State of Florida/United States Department of Education, Office of Civil Rights Partnership Report and Commitments: 2003 Final Report," <www.fldoe.org/BOG/meetings/2004_01_22/OCR_ExecSum.pdf> (July 20, 2004).

Stiefel, Leanna, Amy Ellen Schwartz, Robert Berne, and Colin C. Chellman (2005, in press). "School Finance Court Cases and Disparate Racial Impact: The Contribution of Statistical Analysis in New York," *Education and Urban Studies.*

Stiefel, Leanna, Amy Ellen Schwartz, and Colin C. Chellman (2003). *Education Finance Research Consortium Condition Report: Beyond Black and White: Patterns and Relationships in New York State Test Score Gaps*, Albany, NY: Center for Policy Research, Rockefeller College, University at Albany.

Stiefel, Leanna, Amy Ellen Schwartz, and Dylan Conger (2002). "Immigrant Children in New York City Public Schools." Presented at the New York University's Wagner Graduate School of Public Service Doctoral Colloquium, May.

Steele, Claude and Joshua Aronson (1998). "Stereotype Threat and the Test Performance of Academically Successful African Americans, " in Christopher Jencks and Meredith Phillips, eds., *The Black-White Test Score Gap*, (Washington DC: Brookings Institution): 401-430.

Sternberg, Betty (2004). "Back to School Address: Learning Initiatives for Excellence," Connecticut State Commissioner of Education, address given August 17, 2004, <www.state.ct.us/sde/bck2schaddress04.doc> (August 20, 2004).

Symonds, Kiley (2003). "After the Test: How Schools are Using Data to Close the Achievement Gap." Bay Area School Reform Collaborative, December 2003, <http://basrc.org/research/index.html> (April 11, 2004).

Thernstrom, Abigail and Stephan Thernstrom (2003). *No Excuses: Closing the Racial Gap in Learning* (New York: Simon & Schuster).

Vaughan, Amanda (2001). "Closing the Gap: How the Alabama Reading Initiative Is Transforming Reading Instruction for All Students." <ftp://ftp.alsde.edu/documents/50/ClosingTheGap.pdf> (July 19, 2004).

Wells, Amy Stuart, and Robert Crain (1994). "Perpetuation Theory and the Long-Term Effects of Desegregation." *Review of Educational Research*, v. 64, n. 4 (Winter): 531-555.

Word, E., J. Johnston, H. Bain, B. Fulton, J. Zaharies, M. Lintz, C. Achilles, J. Folger, and C. Breda (1990). "Student/Teacher Achievement Ratio (STAR), Tennessee's K-3 Class Size Study: Final summary report, 1985-1990." Nashville, TN: Tennessee State Department of Education.

## VII. Tables

**Table 1: Distribution of Schools and Students Tested by Racial Accountability Status of School, 2001-02**

| | Schools | | Students | |
|---|---|---|---|---|
| **4th Grade** | Number | Percent | Number | Percent |
| Schools, White Accountable only | 953 | 42.4 | 72,484 | 34.6 |
| Schools, Non-white Accountable only | 523 | 23.3 | 55,503 | 26.5 |
| Schools Accountable for Both Groups | 773 | 34.4 | 81,712 | 39.0 |
| Total | 2,249 | 100.0 | 209,699 | 100.0 |
| **8th Grade** | | | | |
| Schools, White Accountable only | 392 | 36.7 | 42,027 | 20.9 |
| Schools, Non-white Accountable only | 193 | 18.1 | 35,509 | 17.7 |
| Schools Accountable for Both Groups | 483 | 45.2 | 123,409 | 61.4 |
| Total | 1,068 | 100.0 | 200,945 | 100.0 |

Source: Test score data are from New York State School Report Card, 2001-2002, special file provided by the New York State Education Department.
Note: Sample includes all students from schools with ten or more students tested in ELA and in Math, and excludes schools in which more than 40 percent of the students have an IEP.

White Accountable schools have test performance data for ten or more white students tested and fewer than ten non-white students tested. Non-white Accountable schools have test performance data for fewer than ten white students tested and ten or more non-white students tested.
Number of students is based on availability of performance data for the ELA. Numbers would be slightly different if based on Math data.

**Table 2: Teacher and School Characteristics of Study Sample Compared to Averages for New York State and the Districts Containing Our Case Study Schools**

| | **Study Sample (n=23)[1]** | **New York State** | **Downstate Large City District** | **Upstate Small City District** |
|---|---|---|---|---|
| Median years teaching experience | 15.7 | 13.9 | 11.0 | 10.0 |
| Average number of years at current school | 7.4 | NA | NA | NA |
| Percent of teachers permanently certified | 80.0 | 72.1 | 49.3 | 64.1 |
| Expenditures per pupil (district)[2] | NA | $12,387 | $11,474 | $10,243 |
| Median full-time teacher salary[2] | NA | $50,869 | $48,152 | $38,028 |
| Percent of teachers white[3] | 91.3 | 82.2 | 59.0 | 99.0 |

1 Source: Teacher survey, May-June 2004, Institute for Education and Social Policy, New York University. Source for all other data, unless otherwise noted, is the 2003 State of Learning/Chapter 655 report from the New York State Education Department, which contains school year 2001-02 data at the district-level.
2 Expenditure and teacher salary figures are not adjusted for differential costs across districts.
3 Source for school-level teacher race data: Institutional Master File, Fall 2002 (school year 2002-03) from the New York State Education Department.

| **Table 3: Student Interaction** | **NYS Successful Schools** | | **Markow and Scheer (2004)** | |
|---|---|---|---|---|
| | **Disagree or Strongly disagree** | **Agree or Strongly agree** | **Disagree or Strongly disagree/ NA** | **Agree or Strongly Agree** |
| Students from different racial backgrounds frequently hang out together outside of class time. | 5.3 | 94.7 | 28.0 | 72.0 |
| Teachers encourage students of different backgrounds to work together on projects or classwork. | 9.6 | 90.4 | 13.0 | 87.0 |
| Teachers and administrators encourage students of different racial backgrounds to interact together outside of class. | 20.0 | 80.0 | 26.0 | 74.0 |

**Table 4: Which of these two approaches is most important to helping students learn?**

| | **Percent Agreeing** |
|---|---|
| Celebrating differences in student backgrounds | 68.2 |
| Ignoring student background differences as much as possible | 31.8 |

**Table 5: How much do the following factors explain why some student groups have lower test scores than other groups of students? (Percent responding "A lot" or "Some.")**

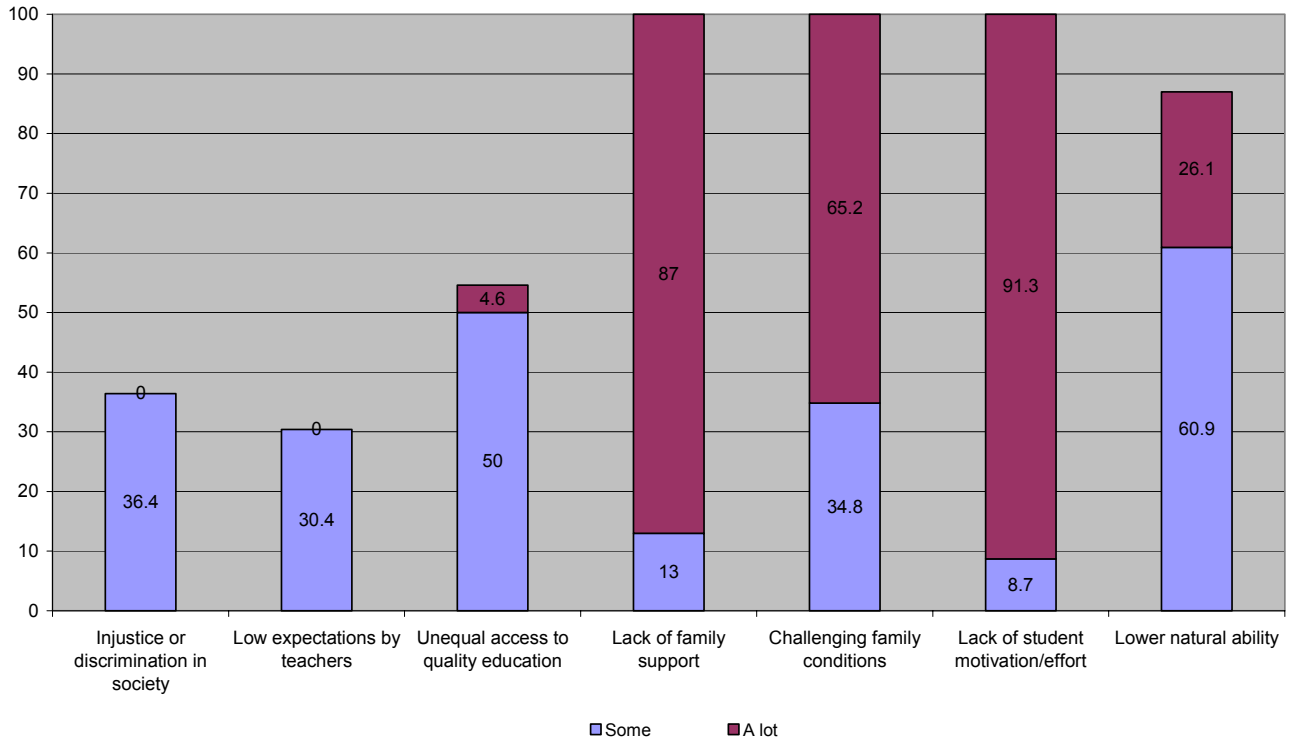
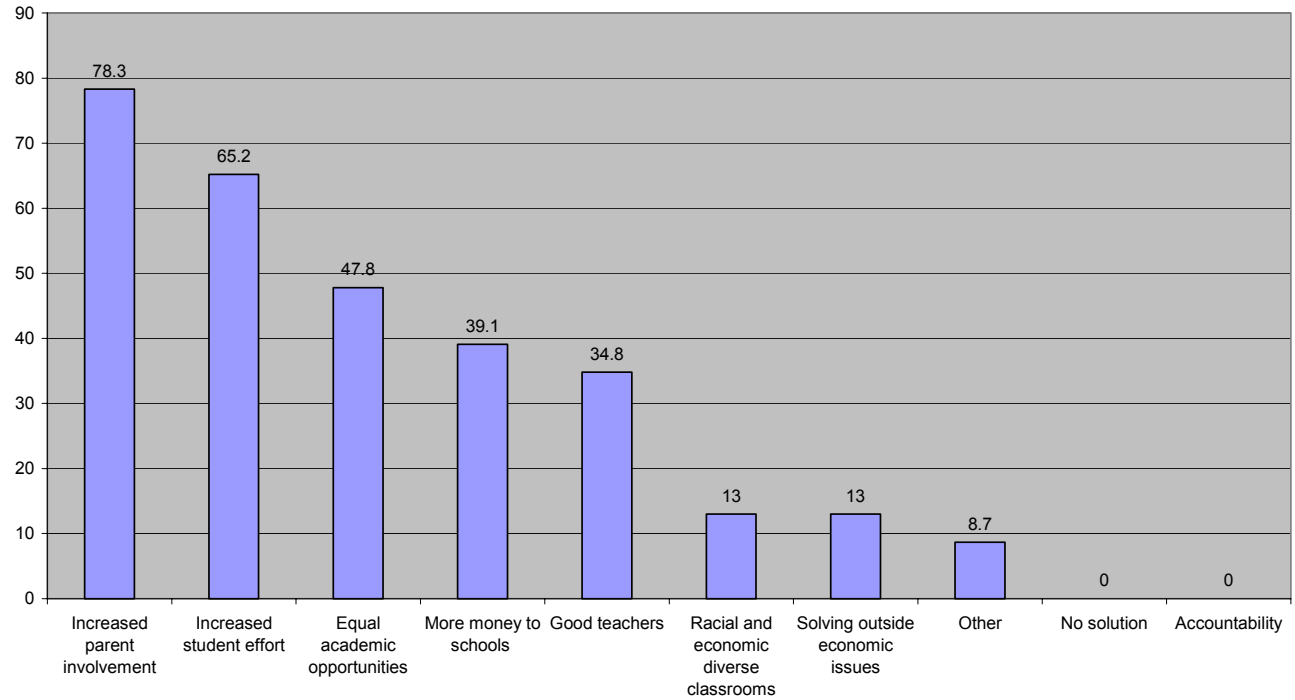



**Table 6: Which three of the following would work best to reduce test score gaps?**



| **Table 7: Principal Leadership** | **Strongly disagree** | **Disagree** | **Agree** | **Strongly agree** |
|---|---|---|---|---|
| Has confidence in the expertise of teachers | 0.0 | 0.0 | 47.8 | 52.2 |
| Takes a personal interest in the professional development of teachers | 0.0 | 9.5 | 47.6 | 42.9 |
| Places the needs of children ahead of personal and political interests | 0.0 | 13.6 | 36.4 | 50.0 |
| Collaborates with teachers and staff to make this school run effectively | 0.0 | 0.0 | 42.9 | 57.1 |
| Understands how children learn | 5.0 | 0.0 | 35.0 | 65.0 |
| Promotes parent and community involvement | 0.0 | 4.5 | 36.4 | 59.1 |
| Is an effective manager who makes the school run smoothly | 4.5 | 4.5 | 31.8 | 59.1 |

| **Table 8: Which of these two approaches is most important in reducing test score gaps?** | **Percent Agreeing** |
|---|---|
| Ensuring that all students are treated equally even if it means some students receive less attention than they need | 23.8 |
| Treating students differently so that their outcomes are similar, even if it means some students receive more attention than others | 76.2 |

**Table 9: How frequently do you….**

| | **Never** | **A few times a year** | **A few times a month** | **A few times a week** |
|---|---|---|---|---|
| Use data to understand the skill gaps of low-achieving students? | 4.5 | 50.0 | 31.8 | 13.6 |
| Administer ongoing assessments to students? | 0.0 | 17.4 | 43.5 | 39.1 |
| Discuss low-performing student achievement data with colleagues? | 0.0 | 13.0 | 65.2 | 21.7 |

**Table 10: Professional Development Activities in the Past Year**

| | Percent |
|---|---|
| Literacy instruction for low-performing students | 69.6 |
| Linking performance data to instructional strategies | 47.8 |
| Math instruction for low-performing students | 39.1 |
| Analyzing low-performing student data | 34.8 |
| Student behavior and discipline | 26.1 |
| Increasing parent/community involvement | 26.1 |
| Teaching heterogeneous groups | 21.7 |
| Teaching English as a Second Language students | 21.7 |
| Multicultural diversity issues | 17.4 |

**Table 11: How much do you agree with the following statements?**

| | Strongly disagree | Disagree | Agree | Strongly agree |
|---|---|---|---|---|
| Teachers regularly communicate with parents about how they can help their children learn | 0.0 | 4.5 | 50.0 | 45.5 |
| Parents understand the standards and expectations teachers hold for students | 0.0 | 36.4 | 59.1 | 4.5 |
| I ask parents for their suggestions about how to work more effectively with their child | 10.5 | 26.3 | 52.6 | 10.5 |

**Table 12: The extent to which you agree or disagree with each of the following:**

| | Strongly disagree | Disagree | Agree | Strongly disagree |
|---|---|---|---|---|
| My school places a priority on eliminating the test score gap | 5.3 | 15.8 | 52.6 | 26.3 |
| My district places a priority on eliminating the test score gap | 5.3 | 21.1 | 42.1 | 31.6 |
| Closing the test score gap has been a primary goal for school leaders | 5.0 | 40.0 | 15.0 | 40.0 |
| School leaders set measurable goal for closing the test score gap | 15.8 | 26.3 | 42.1 | 15.8 |
| School leaders encourage or lead school-wide inquiry into the test score gap | 16.7 | 27.8 | 33.3 | 22.2 |