NBER WORKING PAPER SERIES

IMPROVING EDUCATIONAL OUTCOMES FOR POOR CHILDREN

Brian Jacob
Jens Ludwig

Working Paper 14550
http://www.nber.org/papers/w14550

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2008

A version of this paper was presented at the Institute of Research on Poverty conference "Changing Poverty," which was held at the University of Wisconsin-Madison May 29-30, 2008, with financial support from the Assistant Secretary for Planning and Evaluation of the U.S. Department of Health and Human Services and the Russell Sage Foundation. This paper is forthcoming in fall 2009 in the Russell Sage volume Changing Poverty coedited by Maria Cancian and Sheldon Danziger. Thanks to Helen Ladd, Betsey Stevenson, the editors, and conference participants at the University of Wisconsin's Institute for Research on Poverty and the Philadelphia Federal Reserve Bank and University of Pennsylvania for helpful comments. All opinions and any errors are of course ours alone. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Improving Educational Outcomes for Poor Children
Brian Jacob and Jens Ludwig
NBER Working Paper No. 14550
December 2008
JEL No. I20

ABSTRACT

This review paper, prepared for the forthcoming Russell Sage volume Changing Poverty, considers the ability of different education policies to improve the learning outcomes of low-income children in America. Disagreements on this question stem in part from different beliefs about the problems with our nation's public schools. In our view there is some empirical support for each of the general concerns that have been raised about public schools serving high-poverty student populations, including: the need for more funding for those school inputs where additional spending is likely to pass a benefit-cost test; limited capacity of many schools to substantially improve student learning by improving the quality of instruction on their own; and the need for improved incentives for both teachers and students, and for additional operational flexibility. Evidence suggests that the most productive changes to existing education policies are likely to come from increased investments in early childhood education for poor children, improving the design of the federal No Child Left Behind accountability system, providing educators with incentives to adopt practices with a compelling research base while expanding efforts to develop and identify effective instructional regimes, and continued support and evaluation of a variety of public school choice options.

Brian Jacob
Gerald R. Ford School of Public Policy
University of Michigan
735 South State Street
Ann Arbor, MI 48109
and NBER
bajacob@umich.edu

Jens Ludwig
University of Chicago
1155 East 60th Street
Chicago, IL 60637
and NBER
jludwig@uchicago.edu

I. INTRODUCTION

One of the best ways to avoid being poor as an adult in America is to obtain a good education. As Magnuson and Votruba-Drzal (2009) note, individuals with higher academic achievement and more years of schooling earn more than those with lower levels of human capital. This is not surprising given that we believe that schooling makes people more productive, allowing them to command higher wages in the labor market.

Yet, in modern America disadvantaged children face an elevated risk for a variety of adverse educational outcomes. According to the 2007 National Assessment of Educational Progress (NAEP), only 16 percent of fourth grade students eligible for free lunch score at proficient levels in reading compared with 44 percent of fourth graders whose family incomes are above the eligibility cutoff for free lunch; the disparity in math scores is even larger, 21 versus 53 percent (NCES 2007). Equally large disparities in achievement test scores are observed between whites and minority racial or ethnic groups, with gaps that show up as early as age three or four (Jencks and Phillips, 1998, Neal, 2007).[1]

Understanding *why* children's outcomes vary so dramatically along race and class lines in America is central to formulating effective education policy interventions. Disagreements about how to improve schooling outcomes for poor children stem in part from different beliefs about what problems underlie the unsatisfactory outcomes found in too many of our nation's public schools. Broadly speaking, critics tend to invoke, at least implicitly, one of the following reasons why children in high-poverty schools are not performing as well as we would like:

[1] Note that these disparities in schooling outcomes along race and class lines are not simply due to immigration into the U.S. by those with low initial levels of English or other academic skills, since for example reading and math disparities in NAEP scores are large among 4th and 8th graders even between non-Hispanic whites versus non-Hispanic blacks (www.nces.ed.gov/nationsreportcard/nde, accessed on August 4, 2008).

1. Schools serving poor and minority students have fewer resources (e.g., text books, teachers, support services) than they need to educate the disadvantaged students they serve. In this case, a potential solution would be to provide more money to disadvantaged schools.
2. High-poverty schools lack the capacity to substantially improve student learning, independent of financial resources. Under this perspective the teachers and administrators in highly disadvantaged school districts are thought to lack the skills or knowledge necessary to improve the quality of instruction on their own. Potential solutions to this problem would involve helping schools improve the quality of their standard operating practices, for example by helping implement specific new instructional or organizational practices (i.e., curriculum, instruction, school organization) and/or increasing the instructional capacity of staff in these schools through professional development or more selective hiring. The success of this approach depends critically on whether there are particular instructional, organizational, human resource or other practices that are known to be effective, and how easily these practices could be implemented across different contexts.
3. High-poverty schools do not have sufficient incentives and/or flexibility. High-poverty schools are under-performing because teachers and administrators are not working hard enough, they are not working toward the right goal, or they have good local knowledge about what would work best but they are not able to implement these ideas because of centralized authority. Proponents of this perspective often claim that without clarifying the key objectives of school and holding key actors accountable, additional spending will simply be squandered. Under this view, the solution would be to enhance incentives and provide local actors more flexibility through policies such as school choice or accountability.

4. Schools matter only so much. The real problem rests with the social context in which schools operate – namely, the family, neighborhood and peer environments that low-income children experience. Adopting accountability or market-oriented education reforms without changing social policy more broadly will simply punish educators for factors beyond their control, and potentially drive the most able teachers toward schools serving less disadvantaged students. In this case, a necessary condition for making serious improvements in poor children's schooling outcomes would involve fixing the other social ills associated with poverty that impair children's learning outcomes, such as poor health care or high unemployment.

For some reason current policy debates often seem to be argued as if the problems above are mutually exclusive. In contrast, we believe that there is likely some truth to each of these major explanations. Moreover, it seems plausible that the underlying problems confronting our schools might interact in complicated ways. For example, it may be the case that certain curriculum reforms are effective only if they are accompanied by an increase in resources such as student support services, or by an increase in teacher quality generated by better professional development and/or reforms to hiring and tenure policies

Research over the past four decades has unfortunately fostered the impression that when it comes to interventions for the most disadvantaged public schools in our country "nothing works." One of the first studies to do so was the landmark 1966 report by sociologist James Coleman and his colleagues. By drawing on a large, nationally representative sample, the Coleman Report found that most of the variation in student achievement occurs within rather than across schools, that family background is the strongest predictor of academic achievement, and that most measurable school inputs like student-teacher ratios are only weakly correlated

with student outcomes. Subsequent evaluation studies also tended to be disappointing and helped contribute to a sense of pessimism about the ability of schools to lift poor children's life chances (Levin, 1977, Glazer, 1986, Jencks, 1986).

In contrast, our chapter offers a message of tempered optimism. The past few decades have seen a dramatic improvement in the technology of education policy evaluation, which has enhanced our ability to uncover moderately-sized program impacts within the complex environment that determines schooling outcomes. The available evidence reveals a number of potentially promising ways to improve the learning outcomes of low-income children. This is not to say that everything works: many current and proposed education policies either enjoy no empirical support for their effectiveness, or in some cases have strong empirical evidence for their ineffectiveness. But a careful sifting of the empirical evidence identifies a selected set of interventions that seem to be promising.

Our optimism is tempered by the recognition that even the most successful educational interventions will reduce, but not eliminate, racial and social class disparities in educational outcomes. But the usual standard of public economics for judging policy interventions is that they generate benefits that exceed their cost. Under this benefit-cost standard, education policies that are capable of improving poor children's schooling outcomes by enough to justify the costs of these new programs are worth doing, even if these programs by themselves are not enough to equalize learning opportunities for all children in America.

While we recognize that the social context of children's lives plays a large role in their schooling outcomes, the focus of this chapter involves school-specific policies. For a comprehensive discussion of the direct link between a child's family background and their learning outcomes, see Magnuson and Votruba-Drzal (2009).

The distinction between the role of family background and school-specific policies is blurred by the fact that the concentration of low-income and minority students within schools can have important implications for the success of school-specific policies.[2] A growing body of empirical evidence suggests that policies that can effectively reduce the amount of racial (and by implication socio-economic) segregation within schools may help improve the schooling outcomes of disadvantaged children (see Vigdor and Ludwig, 2008 for a review). Yet there are substantial barriers – both logistical and political – to further integrating schools along race or class lines. There is substantial residential segregation in our country, which limits the amount of school desegregation that can be accomplished since the U.S. Supreme Court made it very difficult to enact across-district plans in the *Milliken v. Bradley* decision (418 US 717, 1974). More generally, the majority of Supreme Court decisions on school desegregation since the early 1970s have restricted the ability of districts to promote integration (Kahlenberg, 2001), including two recent decisions striking down desegregation plans in Seattle and Louisville. It may be that school desegregation efforts are unable dramatically improve the lot of poor children in American schools. In any case in what follows, we assume that the distribution of poor and minority children across schools is fixed, and consider the ability of different education policies to improve their learning outcomes given this reality.

The remainder of the chapter is organized as follows. In Section II, we review three specific areas in which we believe increased resources may yield important benefits for poor children: early childhood interventions, class size reductions in the early grades that hold the quality of teachers constant, and bonuses to attract and retain the highest-quality teachers in disadvantaged schools. In Section III, we examine reforms that do not require a large investment

[2] For example, all else equal the typical teacher seems to prefer working in schools that serve more affluent and less racially diverse student bodies (Hanushek, Kain, and Rivkin, 2004), which may have implications for policy efforts designed to improve the quality of teachers in high-poverty areas.

of resources, but instead seek to make specific changes in the ways schools operate, ranging from changes in curriculum and instruction to changes in teacher hiring and promotion policy. The available evidence for effective changes in school practices are limited, but this problem could be overcome if federal and state governments did more to require schools to use methods that have been proven effective (Jacob and Ludwig, 2005). In Section IV, we review the literature on school accountability and school choice policies, which aim to change the goals, incentives and organizational structure within which schools and families operate. The evidence on school accountability is moderately encouraging, though recent findings suggest that the design of accountability policies matters a great deal. There is less evidence that current public or private school choice plans improve student outcomes, but there are likely to be benefits from continued experimentation with and evaluation of a variety of choice options by states and school districts. Finally in Section V, we summarize what we believe the available empirical evidence suggests are the most promising education policies for improving the life chances of disadvantaged children.

II. SCHOOL RESOURCES

The question of whether "money matters" has been the subject of contentious debate in the research literature for the past 40 years. Isolating the causal effects of extra school funding is complicated by the possibility that compensatory spending may be directed towards those schools serving the most disadvantaged students, and adequately controlling for all aspects of student disadvantage is quite difficult in practice. A few recent studies have used policy changes that generate sharp changes in school funding and find some impact on student outcomes (Figlio, 1997, Card and Payne, 2002, Guryan, 2003). Overall, however, there is fairly weak support for the idea that increases in unrestricted school funding will improve student outcomes (see for

example Hanushek 1997). In contrast, there is stronger evidence that some targeted increases in specific school inputs can improve student outcomes. In this section, we discuss three specific areas in which we believe increased resources may yield important benefits for poor children: increased investments in early childhood education; class size reductions in the early grades; and targeted salary bonuses to help disadvantaged schools recruit and retain better teachers.

A. Early childhood education

Disparities in academic achievement by race and class are apparent as early as age three and four, well before children enter kindergarten. Recent research in neuroscience, developmental psychology, economics, and other fields suggests that the earliest years of life are a particularly promising time to intervene in the lives of low-income children (Shonkoff and Phillips, 2000, Carniero and Heckman, 2003, Knudsen et al., 2006). Studies show that early childhood educational programs can generate learning gains in the short-run and, in some cases, improve the long-run life chances of poor children. Moreover, the benefits generated by these programs are large enough to justify their costs.

The Perry Preschool and Abecedarian programs are commonly cited as examples of how high-quality preschool services can change the lives of low-income children. A small group of children who participated in the these programs in the 1960s and 1970s have been followed for many years and on average have better outcomes in a range of domains compared to a randomly assigned group of control children (Schweinhart et al., 2005; Ramey and Campbell, 1979; Campbell et al., 2002; Barnett and Masse, 2007). Despite the high cost of these programs, these studies suggest that the total economic benefit of these programs exceeded their costs (Belfield et al., 2006; Barnett and Masse, 2007). While these results are encouraging, it is important to keep

in mind that these are model programs that were unusually intensive and involved small numbers of children in just two sites.

Yet evidence on publicly funded early education programs, which illustrate what can be achieved for large numbers of children in programs of variable quality, is also very encouraging. A recent random-assignment evaluation of Head Start found positive short-term effects of program participation on a variety of cognitive skills on the order of .2 to .4 standard deviations,[3] with typically positive effects on non-cognitive outcomes as well that are usually not statistically significant. A rigorous evaluation of Early Head Start, a program serving children under age three in a mix of home and center-based programs, found positive effects on some aspects of parent practices and children's development, but the effects were generally smaller than for Head Start (Love et al. 2002).

Wong et al. (2008) find that several recent state-initiated universal pre-K programs have even larger impacts on short-term cognitive test scores than does Head Start (see also Gormley et al., 2005; Gormley and Gayer, 2005). Why might the new state pre-K programs generate larger cognitive gains than does Head Start? One explanation is that pre-K programs hire more qualified teachers, pay them more, and offer a more academically oriented curriculum than do Head Start programs. Another explanation is that the Head Start comparison group received more center-based care than did children in the pre-K comparison group.[4] A third possible explanation is that the recent Head Start study relies on a rigorous randomized experimental

[3] The official evaluation of the program found impacts on elementary pre-reading and pre-writing skills equal to about 0.3 and 0.2 of a standard deviation respectively (Puma et al. 2005; Ludwig and Phillips, 2007). If one calculates Head Start impacts by pooling together the three- and four-year-olds in the experiment, the increased statistical power leads to significant program impacts on math scores and on almost all of the other main cognitive skill outcomes in the report (Ludwig and Phillips, 2007).

[4] http://www.northwestern.edu/ipr/events/briefingdec06-cook/slide16.html

design, while the research design of the new state pre-K studies is more susceptible to bias from selection problems (see Ludwig and Phillips, 2007). [5]

While these short-run achievement gains from both Head Start and newer state pre-K programs are impressive, the crucial question is whether these effects persist over time. In order to explore longer-run impacts, one must rely on non-experimental studies of children who participated in Head Start decades ago and control for potential confounding factors (Currie and Thomas, 1995; Garces, Thomas and Currie, 2002; Ludwig and Miller, 2007; Deming, 2007). These studies suggest lasting effects on schooling attainment and perhaps criminal activity, although test score effects appear to fade out over time. Nonetheless, the positive long-term effects seem large enough to generate benefits that likely outweigh the program costs.

It is possible that the long-term effects of Head Start on more recent cohorts of children may be different from those for previous cohorts of program participants because program quality changes over time, or because the developmental quality of the environments that children would experience as an alternative to Head Start change. But the short-term test score impacts that have been estimated for recent cohorts of Head Start participants appear to be similar to what we see for earlier cohorts of children for whom we also now have evidence of long-term benefits. So there is room for cautious optimism that Head Start might improve the long-term outcomes for recent waves of program participants as well, even though this cannot be directly tested for many years (Ludwig and Phillips, 2007).

Although preschool interventions represent a promising way to improve the life chances of poor children, their success is not well reflected in federal government budget priorities, which allocate nearly seven times as much money per capita for K-12 schooling as for pre-

[5] Some prior research has found that while pre-kindergarten programs improve cognitive outcomes, they may have adverse effects on non-cognitive outcomes like self-control (see, for example, Magnuson, Ruhm and Waldfogel, 2004). The evidence on this issue is not yet available for the new state pre-K programs.

kindergarten (pre-K), other forms of early education and child care subsidies for three- to five-year-olds. Most social policies attempt to make up for the disadvantages poor children experience early in life. But given the substantial disparities between poor and non-poor children that already exist among very young children, it is perhaps not surprising that many disadvantaged children never catch up. There may be very high returns indeed to increased investments in improving young children's school readiness with proven, high-quality programs.

B. Class size reduction

Reducing average class sizes may enable teachers to spend more time working with individual students, tailor instruction to match children's needs, and make it easier for teachers to monitor classroom behavior. Class size reductions are not inexpensive, however. They require hiring additional teachers and in some cases expanding a school's physical space. Nonetheless, the best available evidence suggests that class size reduction, holding teacher quality constant, can improve student outcomes by enough to justify these additional expenditures, with benefits that are particularly pronounced for low-income and minority children.

The evidence in favor of class size reduction comes mainly from Tennessee's Project STAR, which randomly assigned 11,600 students in grades K-13 and 1,330 teachers to small classes (13-17 students), regular size classrooms (22-25 students), or regular size classrooms that also included a teacher's aide.[6] Teachers were also randomly assigned to different classroom environments in this study, which ensured that the average quality of teachers in small versus regular size classrooms is the same, a critical point to which we will return within the context of policy efforts to take class size reduction to scale. Class size reductions of around one-third during these early grades increased Stanford-9 reading and math scores by around 0.12 standard deviations for whites and 0.24 standard deviations for blacks. Impacts were larger for students

[6] This section is based on the summary of Project STAR research by Schanzenbach (2006/7).

attending mostly black schools, although even within such schools, black students benefited somewhat more than did whites (Krueger and Whitmore, 2002). Similarly, impacts were somewhat larger for students eligible for the free lunch program.

Follow-up evaluations of STAR find long-term benefits of attending a small elementary school class – the test score impacts seem to persist through 8th grade, although they decline by one-half to two-thirds (Achilles et al., 1993, Nye et al., 1995, Krueger and Whitmore, 2001). As with the short-term effects, the eighth grade test scores reveal larger gains among low-income and minority students. Moreover, black students in the treatment group were roughly 5 percentage points (or 15 percent) more likely to take a college entrance exam (i.e., the SAT or ACT) during high school (Krueger and Whitmore, 2001, 2002).[7] The test score gains induced by smaller classes in STAR are probably large enough to justify the costs, even when we focus just on the benefits that arise from test scores on future earnings alone (Krueger, 2003, Krueger and Whitmore, 2001, Schanzenbach, 2006/7).

The results from Project STAR are encouraging, but they come from a controlled experiment that held the quality of teachers constant by randomly assigning teachers as well as students across classrooms. A sobering example of the challenges of taking class size reduction to scale comes from the California experience. California introduced a statewide initiative to reduce primary grade class sizes in 1998-99, which required schools to not only hire a large number of new teachers, but also find additional physical space for the new classrooms (Borhnstedt and Stecher 2002; Jepsen and Rivkin 2002). The policy was implemented over a short time period. Many low-income school districts found it difficult to hire qualified teachers, and had trouble arranging for adequate classroom space. Jepsen and Rivkin (2002) argue that

[7] Effects for outcomes such as criminal involvement or teen births yield point estimates that are in the direction of beneficial CSR impacts, but are imprecisely estimated (Schanzenbach, 2007).

much of the benefit from smaller classes in California was lost due to reductions in average teacher quality, particularly in lower-income urban school districts.

California's experience suggests the complexity of taking education policy reforms to a large scale. It suggests the potential value of focusing class size reductions in low-income districts or schools, as well as the importance of careful implementation of the program so as to minimize any adverse effects on teacher quality or physical capital. So long as these implementation efforts did not raise costs too much relative to STAR, it still seems plausible that a large scale program to reduce class sizes might pass a benefit-cost test.

C. Bonuses for teaching in high-needs schools or subjects

Research has identified substantial variation across teachers in the ability to raise student achievement. These studies estimate teacher effectiveness by comparing changes in student achievement scores across classrooms, controlling for student, classroom and school characteristics that influence achievement regardless of the teacher. Because these studies attempt to isolate the value that a teacher adds to student achievement, they are referred to as "teacher value-added" studies. These studies document substantial variation in teacher effectiveness, both within and across schools. According to a recent analysis of NYC elementary school math teachers, for example, students whose teacher falls in the top quarter of effectiveness learn roughly 0.33 test score standard deviations more in a single year than students whose teachers are in the bottom quarter (Kane et al. 2006).[8] If disadvantaged children were taught by the most effective teachers, disparities in schooling outcomes might be narrowed.

[8] There are three concerns with value-added measures of effectiveness. One involves the statistical precision of the measures; the other involves the possible biases inherent in such measures. First, given the small number of students that an individual teacher works with in any given year (i.e., 15 or 20 for many elementary school teachers) and the imperfect reliability of student achievement tests, teacher "value-added" measures are generally measured with considerable error (Kane and Staiger, 2002a,b). That is, the "confidence interval" around an estimate of a teacher's effectiveness is often quite large. A second problem is that students are not randomly assigned to

Value-added measures of teacher effectiveness are not very strongly correlated with the easiest-to-observe characteristics of teachers. On the one hand, novice teachers are less effective than more experienced ones, but this experience premium disappears after the first few years of teaching (Rockoff 2004). Teachers who have higher scores on the SAT or various teaching exams are generally more effective than others (Harris and Sass 2007, Clotfelter et al. 2007). On the other hand, there is little if any difference in the achievement gains of students assigned to traditionally certified, alternatively certified or uncertified teachers (Kane et al. 2006; Boyd et al. 2005). Similarly, researchers have found that teachers who have advanced degrees are not more effective than other teachers (Hanushek 1997).

Hence, the policy challenge in this domain is to induce more effective teachers to teach in schools serving the most disadvantaged children, knowing that effectiveness cannot easily be measured. Given the relationship between wages and teacher labor supply (see, for example, Hanushek, Kain, & Rivkin, 2004; Stinebrickner 2002; and Scafidi, Stinebrickner, & Sjoquist, 2007), one strategy is to entice teachers to work in hard-to-staff schools and/or subjects through financial incentives such as targeted salary increases or bonuses since. While many states have adopted some sort of financial assistance programs (e.g., loan forgiveness, mortgage assistance, salary supplements), there has been little systematic evaluation (Imazeki 2007, Gaurino et al. 2006, Glazerman et al. 2006).

Evaluation of a targeted teacher bonus in North Carolina suggests the program was successful in reducing teacher turnover rates in low-income or low-performing schools (Clotfelter, Glennie, Ladd and Vigdor, 2006). In California from 2000 to 2002 the Governor's

---

classrooms, and hard-to-measure student characteristics may influence the rate of growth in children's test scores not just their levels. So value-added measures may partly confound the causal contribution of the teacher to student learning with those of the characteristics of their students (Rothstein 2007). A third concern is that the impact of having an effective teacher may fade out over time (Jacob et al. 2008).

Teaching Fellowship program offered academically talented teaching candidates $20,000 bonuses to work in low-performing schools, which was found to increase the rate at which such teachers started working in disadvantaged schools but did not seem to influence subsequent retention rates (Steele et al. 2008). Key questions that remain include whether such programs will successfully lure new teachers into the profession who would not otherwise have taught, rather than just re-sort a fixed pool of teachers across schools, and how long any talented new teachers who are drawn into teaching will remain. Moreover, financial incentives that are not tied to teacher performance might induce ineffective teachers to locate in hard-to-staff schools.

III. CHANGING SCHOOL PRACTICES

In the previous section, we identified several reforms that seem likely to improve achievement of poor children. Yet all of these policy changes are expensive, and so must compete for scarce government resources with a variety of other pressing societal concerns. Some observers of America's schooling system remain skeptical that additional spending is needed to improve the learning outcomes of poor children. They argue that by improving the ways in which schools are organized, including the way they deliver instruction, one could improve student achievement with few additional resources. This line of reasoning assumes there is good evidence on which practices are most effective, but that school personnel do not have the capacity to identify or implement these programs on their own.

In the section below, we review the evidence on a variety of reforms, and find some low-cost changes in school operating practices that seem to improve student outcomes, including changes to school organization, classroom instruction and teacher hiring and promotion. What remains unclear is the degree to which these "best practices" have not been more widely adopted

because of a lack of information, political resistance, bureaucratic inertia or some combination of these (or other) barriers.

A. Curricular and Instructional Interventions

In 2002, the Institute for Education Sciences (IES) within the U.S. Department of Education created the What Works Clearinghouse (WWC) in order to collect and disseminate scientific evidence on various educational interventions. A brief review of the WWC website highlights the lack of convincing evidence on curricular interventions. For example, WWC found only one program with strong evidence of improving reading achievement among beginning readers (i.e., Reading Recovery), one program for elementary school math (i.e., Everyday Mathematics) and two programs for middle school math (I Can Learn Algebra/Pre-Algebra and Saxon Middle School Math). Moreover, WWC excluded many studies that did not meet a minimal standard of evidence, and even the programs for which there is strong evidence of success according to WWC show a disappointingly low level of rigorous research. For example, the Everyday Mathematics curriculum was introduced in 1983 and has been used in 175,000 classrooms by around 2.8 million students (WWC 2007). Yet, WWC only found four studies of Everyday Mathematics that met their minimal standards of evidence "with reservations." This lack of success is not limited to traditional "paper and pencil" curricula. There is also little evidence that educational software programs have significant impacts on student learning despite their growing popularity (Dynarski et al. 2007).

A more recent approach to school improvement known as Comprehensive or Whole School Reform (CSR), attempts to improve many different aspects of the school at the same time, hopefully in a fashion that makes the changes complementary and reinforcing. For example, a CSR model may combine curriculum materials, professional development, teacher

mentoring, reorganization of the school day (e.g., block scheduling) and school structure (e.g., schools-within-a-school). Examples include Success for All (Borman and Hewes 2003), Comer Schools (Cook et al. 2000), Direct Instruction (CRSQ 2006), Accelerated Schools (Bloom et al. 2001), America's Choice (CRSQ 2006), Career Academies (Kemple and Scott-Clayton 2004), Project GRAD (Snipes et al 2006), First Things First (Quint et al 2005) and Talent Development (Kemple et al 2005). Unfortunately the evaluation evidence about the effectiveness of CSR programs is, like with curricular reforms, quite limited.

At the elementary school level, a few models have been shown to improve student outcomes. One of the more promising interventions seems to be Success for All (SFA), a comprehensive whole-school reform model that operates in more than 1,200 mostly high-poverty Title I schools.[9] SFA focuses on reading, with an emphasis on prevention and early intervention. Children receive 90 minutes each day of reading instruction in groups that are organized across grade levels based on each child's current reading level, which helps teachers to target instruction. Students engage in cooperative learning exercises in which they discuss stories or learn from each other, which helps reinforce what teachers do and build social skills. Children are assessed at eight-week intervals, using both formal measures of reading competency and teacher observations. Children who fall behind are given extra tutoring or help with other problems, such as health or behavior problems. The program utilizes regular classroom teachers who receive brief initial training, ongoing coaching, and other forms of support and professional development.

[9] Several other elementary school models show promise, including Direct Instruction (CRSQ 2006). One of the only other reform models that has been rigorously evaluated is the Comer Schools program, although one problem identified by the research is the limited degree of difference between Comer treatment schools and control schools in the implementation of Comer-style school practices (Cook et al. 1999, 2000).

A random assignment evaluation of SFA documented that at the end of three years, students in the treatment schools scored roughly 0.2 standard deviations higher than students in the control schools on a standardized reading assessment. These effects were equivalent to about one-fifth the gap between low and high socioeconomic-status children (Borman et al., 2007). SFA costs about $950 per student per year (Borman and Hewes 2003), with about two-thirds of this cost associated with the program's tutoring component. Current spending under the federal government's Title I program is around $880 per eligible student, and that there is little evidence that Title I funding as typically deployed by public schools translates into much gain in student learning (Gordon, 2004, van der Klaauw, 2008). One implication is that in schools serving predominantly poor students, SFA could be implemented mostly by just redeploying existing Title I funds. If the experimental evaluation of SFA is correct and the program's impact is anything like 0.2 standard deviations, this type of funding shift would easily pass a benefit-cost test, since Krueger (2003) shows that test score increases of this magnitude would increase the present value of each student's lifetime earnings profile by thousands of dollars.

At the high school level, CSR models incorporate similar features, often involving a reorganization of the larger high school into smaller learning communities, referred to as "small schools," "schools within a school," and "learning academies." Small learning communities aim to provide a more personalized learning environment that will better engage and motivate students, and prevent at-risk students from "falling through the cracks." Other common reform features include work-based learning opportunities (e.g., internships) and assistance for students who enter high school with poor academic skills (Herlihy and Quint 2006).

However to date there is little evidence that, as currently implemented, these high school CSR approaches improve student outcomes. The most rigorous evaluation of a high school CSR

is the random assignment study of Career Academies, which are organized as small learning communities of between 150 and 200 high school students (often housed within larger, comprehensive high schools) that focus on a specific occupation or industry. Unlike the vocational education programs that were popular in the 1970s and 1980s, career academies combine academic and technical curriculum, often using the technical curricula relevant to a particular industry to motivate students. One hallmark of these programs is to establish a connection with employers in the given field, which occurs through mentors and internships of students with local employers.

Early results of the Career Academy evaluation indicated that students in the program reported that their school provided a more personalized learning environment relative to students in the control group. The program raised attendance and completion of course credits in the early years of high school. A later follow-up found no differences in high school completion rates between program and control groups, but did find a substantial impact on earnings among high-risk students (Kemple and Willner, 2008). More research is needed to assess this approach, including evidence about whether the positive effects documented in the prior research would generalize to students who do not volunteer for the program, and whether the career academies studied as part of this evaluation are of higher quality than the typical career academy.

The federal government could do more to improve the quality of research and development of curricular improvements or CRS, by for example holding "design competitions." Federal, state and local governments could provide curriculum developers with incentives to do more rigorous evaluation of their products by requiring that schools devote resources to only those products that have been proven to be effective in randomized trials (see Jacob and Ludwig 2005). The Institute for Education Sciences is moving in this direction. But the FDA's

requirements that new drugs be subject to at least three phases of clinical trials highlights how far we are from providing parents the same assurances about the value of their children's learning experiences at school.

B. Teacher Labor Markets

A key policy challenge for school districts is to induce more effective teachers to teach in high-poverty schools. As discussed above, one strategy is to offer targeted financial incentives to individuals teaching in hard-to-staff schools and/or subjects. But there are also a variety of potential inefficiencies in the way that schools hire, promote, and dismiss teachers. At least some of the problems might be addressed without substantial increase in resources.[10]

One promising approach is to promote alternative pathways into teaching. Traditional certification requirements impose a high cost (both in terms of money and time) on individuals interested in teaching, particularly on those with the best outside labor market options. As a result certification requirements may help dissuade some highly skilled people from entering the teaching profession. Many studies have explored the relative effectiveness of teachers with traditional versus alternative (or no) certification. The emerging consensus is that differences between the groups are relatively small, and that in certain grades and subjects, teachers with alternative certification may actually outperform those with traditional certification (Boyd et al. 2005; Kane, Rockoff and Staiger 2006, Glazerman et al. 2006). Hence, alternative certification may improve the supply of teachers without reducing quality.

Given the wide variation in effectiveness, policies that help school officials identify and hire the best applicants could have an important impact on student outcomes. Unfortunately,

[10] There has been little research on teacher "demand" policies. One reason is the perception that disadvantaged school districts are in a state of perpetual shortage, and thus hire anyone that walks through the door. In reality, while there are often shortages in certain subjects and grade levels, many disadvantaged districts have an ample supply of teachers for most positions. For example, the Chicago Public Schools regularly receives 10 applications for each position.

there is little guidance that researchers can provide in this area. For example, staffing rules, late state budgets and inefficient human resource procedures dissuade many qualified applicants from teaching in urban schools (Levin and Quinn 2003; Levin et al. 2005). This suggests that changes to collective bargaining agreements that reform staffing rules and a more efficient, customer-focused approach to HR would improve teacher quality.[11]

Whatever system is used to hire teachers, it is inevitable that some teachers will not perform well in the classroom. Recognizing that the hiring process is imperfect, virtually all school systems place new teachers on probation for several years, subjecting them to an up-or-out tenure review. However, in practice, public schools typically do not take advantage of the probationary period to obtain additional information about teacher effectiveness and weed out lower-quality teachers.[12] In New York City, for example, only about 50 out of roughly 75,000 teachers were dismissed for performance-related reasons in recent years.[13] In the Chicago Public Schools, only 15 out of 11,621 teachers who were evaluated in 2007 received a rating of unsatisfactory, and only 641 out of 11,621 (roughly 5.5 percent) received a rating of satisfactory. The remaining teachers were rated excellent or superior.[14]

One possible solution is to raise the tenure bar for new teachers, and to deny tenure to those who are not effective at raising student achievement. Some have suggested that this type of evaluation should be based at least in part on teacher value-added scores (e.g., Gordon et al. 2006). Other criteria could also be used in making high-stakes decisions such as teacher dismissals, given the concerns about the reliability and validity of value-added measures

[11] The discussion draws on Jacob (2007). Professional development, including mentoring for novice teachers, is another strategy for enhancing teacher ability. Such policies are complements to the policies discussed in this section. For a review, see Hill (2007).
[12] As mentioned earlier, professional development is clearly an important complement to the dismissal of underperforming probationary teachers.
[13] Personal communication with Jonah Rockoff, February 19, 2008.
[14] Calculations by Brian Jacob from CPS administrative records.

discussed earlier. For example, while commonly measured teacher characteristics have little correlation with student achievement, principals do recognize which teachers are most effective. Jacob and Lefgren (2007) compare principal ratings of teacher effectiveness with "objective" measures of teacher effectiveness calculated using student achievement gains. They find that principals can identify the best and worst teachers, but have little ability to distinguish between teachers in the middle of the ability distribution. This suggests that principal evaluations could also be included as one factor in teacher tenure ratings, both because they may add additional information beyond student test scores, but also because they reduce potential negative effects of relying solely on an output-based measure (e.g., teachers cheating, or teaching narrowly to the test in order to maximize short-run student test scores at the cost of skills that will maximize long-term learning).

One might still be concerned that principals would be hesitant to deny tenure to many teachers. Several years ago, the Chicago Public Schools and Chicago Teachers' Union signed a collective bargaining agreement which gave principals considerable latitude to dismiss untenured teachers. Jacob (2008) finds that the teachers who were selected for dismissal did have more absences and lower value-added scores than other probationary teachers in the same school who were not selected for dismissal, suggesting that principals did consider productivity in making their decisions. However, Jacob (2008) reports that 30-40 percent of principals – including many at very low-performing schools – did not dismiss *any* teachers, suggesting that some administrators may be reluctant to remove poorly performing employees. One reason may be that it is difficult to hire high-quality replacements, and so principals might be reluctant to risk the chance of getting an even worse replacement. It is also possible that principals are reluctant to incur the social and political costs associated with dismissal. In either case, any policy to raise

the tenure bar will inevitably have to grapple with incorporating some system to ensure that difficult decisions are made.

IV. INCENTIVES AND ACCOUNTABILITY

Class size reduction is an "input-based" educational intervention, based on the assumption that schools will perform better with additional resources. Comprehensive school reform is based on the assumption that schools are not utilizing optimal pedagogical practices, and therefore seeks to improve schooling outcomes by prescribing a more effective instructional approach. Both strategies assume that educators are willing to work as hard as they can given resource constraints. Prescriptive approaches that push schools to adopt specific practices also assume that it does not make sense for individuals to continually "reinvent the wheel" and thus relies on centralized decision makers who presumably have better information about "best practices" for any given school than local principals and teachers. An alternative approach to school reform focuses on enhancing both the incentives and flexibility enjoyed by school personnel. While the theories underlying school choice and school accountability differ in important ways, both strategies rely on the core notions of incentives and flexibility. In this section, we review the existing empirical work on how accountability and choice reforms impact student outcomes.

The available evidence to date is probably strongest on behalf of the ability of school accountability systems to change the behavior of teachers and principals, although one lesson from that body of research is the great importance of getting the design of such policies right.

A. Teacher merit pay

Teacher compensation often arises in discussions of school reform. Most public school teachers are paid according to strict formulas that incorporate years of service and credits of

continuing education including Master's and doctorate degrees, despite the fact that research consistently finds that advanced degrees are not associated with better student performance and experience only matters in the first few years of teaching. For this reason, reformers have suggested that a teacher's compensation should be tied directly to her productivity as measured by student performance or supervisor evaluation. Proponents of "pay-for-performance," also known as "merit" or "incentive" pay, argue that it would not only provide incentives for current teachers to work "harder" or "smarter," but also could alter the type of people who enter the teaching force and then choose to remain.

Critics of merit pay emphasize instead that teaching is a collaborative venture and that incentive pay for individuals could harm the teamwork necessary for effective schools. They note that it is difficult to monitor teachers, and that because there are multiple objectives that we would like teachers to accomplish, a system that focuses on one easily observable outcome such as student test scores will likely distort teacher behavior in ways that harm other outcomes. For example, teachers may focus attention more on the subjects that "count" under the performance system, or they may neglect students who they view as too far behind (or even too far ahead). Proponents counter that pay-for-performance systems can focus on teams of teachers or schools, and that one might incorporate principal evaluations as well as test scores to mitigate distortions.

As Murnane and Cohen (1986) note, incentive pay has a long history in American education, though few systems that directly reward teachers on the basis of student performance have survived. While many teachers indicate that their compensation incorporates some aspect of incentive pay, there are few examples of pay that is tied very closely with student performance. One prominent example is the Teacher Advancement Program (TAP), which incorporates incentive pay along with pay for additional professional development activities and

other service. This program is more popular than many "old style" merit pay programs, and there is some evidence that they may improve student performance on standardized tests (Springer et al. 2008). Given this tentative but positive evidence, there could be some benefit from continued experimentation by schools and districts with pay for performance, particularly if this experimentation is subject to careful evaluation as well.

B. School accountability systems

Recent studies suggest that accountability reforms can foster positive changes in behavior by school administrators, teachers, and students. At the same time research provides some warnings that incentive-based reforms often generate unintended negative consequences, such as teachers neglecting certain students, cutting corners or even cheating to artificially raise student test scores. The fact that actors within the school system do respond to changes in incentives highlights both the promise and pitfalls of accountability reform, and underscores the importance of the specific design details of accountability policies.

While studies of school-based accountability policies in the eighties and nineties found mixed results (Jacob 2005), two major accountability reforms from the late 1990s – in Chicago and Florida – have demonstrated positive results. In 1996, the Chicago Public Schools (CPS) introduced a comprehensive accountability policy designed to raise academic achievement. The first component of the policy focused on holding students accountable for learning, by ending a practice commonly known as "social promotion" whereby students are advanced to the next grade regardless of ability or achievement level. In addition, the CPS also instituted a policy that placed low-performing schools on probation, which entailed a modest amount of additional resources along with enhanced monitoring and the threat of future closure.

Jacob (2005) finds that math and reading achievement increased sharply following the introduction of this accountability policy, in comparison to both prior achievement trends in the district and to changes experienced by other large, urban districts in the mid-west. However, for younger students, the policy did not increase performance on a state-administered, low-stakes exam. His analysis suggests that the observed achievement gains were driven by increases in test-specific skills and student effort. This finding is consistent with prior work suggesting that test preparation associated with high-stakes testing may artificially inflate achievement, producing gains that are not generalizable to other exams (see for example Koretz et al. 1991, Koretz and Barron 1998). Note that the policy did lead to substantial achievement gains for older students, and that for older students the test-specific gains did likely represent real learning of the skills found on the high-stakes tests.

Figlio and Rouse (2006) study Florida's A+ Plan for Education, a school-based accountability system that was implemented several years before the introduction of No Child Left Behing. In 1999, schools received "grades" based on their students' test scores on the Florida Comprehensive Achievement Test (FCAT). Under the A+ Plan, schools that received a failing grade (i.e., "F") in two consecutive years were required to offer their students vouchers to enroll in private schools. Schools that faced declining enrollments were subject to staff cuts, and additional sanctions. It was hoped that the stigma associated with the highly visible failing grade, coupled with the threat of competition induced by the vouchers, would provide low-performing schools with an incentive to improve.

Figlio and Rouse (2006) find that the designation of a school as failing led to a significant increase in student math performance, on the order of 5 scale score points, roughly 0.1 standard deviations. The impact of receiving an "F" on a low-stakes math test is about half as large.

Researchers studying later versions of Florida's accountability system find similar effects (Chiang 2007, Rouse et al. 2007). Moreover, they find that much of these learning gains were sustained in the following few years.

While these studies suggest accountability policies can improve student achievement, other studies show that educators respond strategically to test-based accountability in many unintended ways, some of which may have negative consequences for students. Jacob (2005) finds that educators in Chicago responded to the accountability program by placing a larger fraction of low-performing students in special education (thus removing them from the test-taking pool[15]), retaining a larger fraction of students before they reached the grades where they were subject to accountability mandates (i.e., in kindergarten through second grade), and by shifting attention away from subjects such as science and social studies that were not used to determine student or school sanctions under the accountability policy. Each of these responses increased a school's measured performance, even though they seem inappropriate and detrimental to student learning. However, it is also possible that some, and perhaps most, educators viewed these steps as appropriate responses, which would benefit students.[16] A growing body of research also suggests that accountability systems generate larger improvements among students who are relatively closer to the passing threshold used to reward schools (Chakrabarty, 2006; Neal and Schatzenbach, 2007). Finally, Jacob and Levitt (2003) find that the prevalence of teacher cheating rose sharply in low-achieving classrooms following the introduction of Chicago's accountability policy.

[15] Unlike more recent accountability reforms, the program in Chicago did not explicitly monitor the fraction of students who were tested so that placing students in special education would benefit a school rating.

[16] A struggling student may benefit from placement in special education services. Similarly, teachers and administrators often believe that holding children back in an earlier grade provides them additional opportunities to mature and to master basic skills before moving on to higher grades. Also, the focus on math and reading relative to science and social studies was one intended goal of the program.

Two recent studies utilize data from the National Assessment of Educational Progress (NAEP) to assess the impact of standards-based accountability across a variety of states, rather than focusing on case studies of a single jurisdiction's policy. These studies are important because the NAEP data was not used by any state as part of its accountability program, which means educators had no incentive to manipulate test scores in the way that prior studies document has occurred for high-stakes exams in individual states. Both studies find positive achievement effects overall, although they do not find evidence that accountability policies have consistently reduced the racial achievement gap (Carnoy and Loeb 2003; Hanushek and Raymond 2005).

A review of simple national time trends in the recent NAEP data suggests that NCLB may have improved student achievement, particularly the math performance of younger children. However, to our knowledge, there has not been any systematic investigation of the impact of NCLB at a national level which attempts to account for prior achievement trends or the presence of other policies.

Even without evaluation evidence of NCLB directly, the available accountability research suggests a number of modifications to NCLB that could potentially increase the effectiveness of the legislation. The first of these is the adoption of a single achievement standard for all districts in the country. The provision that allowed states to choose the tests they use for measuring student performance has resulted in vast disparities in the academic rigor of accountability requirements across states. Given the political difficulty in imposing a uniform national standard, some national associations of state officials and educators have worked to voluntarily adopt common standards. A second feature of NCLB that has come under scrutiny is the use of a single proficiency level --that is, holding schools accountable for the share of students with

scores above some single cutoff value – since this provides students an incentive to neglect students who are far above or below this threshold. An alternative approach is for schools to be held accountable for the *gains* that all students make each year. A third change to NCLB that has been discussed is to provide states and districts with flexibility to focus on the schools most in need of improvement, particularly if the current level of federal funding for NCLB is not increased substantially.

C. School Choice

Another way to clarify goals or change incentives is to provide parents greater choice of schools for their children through public magnet schools, charter schools, or vouchers for students to attend private schools. Choice proponents suggest that by creating a marketplace where parents can select schools, a choice-based system might generate competition among schools that would improve the quality of schools throughout the marketplace. This theory rests on several assumptions, including that the degree of choice will be large enough to generate meaningful competition. For example, a handful of charter schools with limited enrollment capacity are unlikely to generate meaningful competition in a large district. This suggests that a choice system must permit relatively easy entry into the market by potential suppliers, which includes individuals and organizations wishing to open schools. There must also be easy "exit" from the market that allows (and, indeed, forces) unsuccessful schools to close.[17]

The second set of assumptions involves the information available to parents and the preferences they have for their children's education. Parents must have sufficient information to make an informed choice. Data on school performance must be transparent, accessible and easily understood by parents with varying degrees of sophistication. The effects of choice will

[17] If the administrators and teachers in a public school that loses half of its students to a nearby charter school continue teaching the smaller group of students, or are merely reassigned to other schools in the district, they may not change their practices despite the pressure exerted by the nearby charter.

also depend in large part on the nature of the preferences themselves. Expanded choice is only likely to increase student academic outcomes if achievement is a central concern for parents.

The most rigorous studies to assess parental preferences typically find that parents place a high value on proximity and student racial composition (Glazerman 2000, Calvo 2006, Hastings et al. 2006). On average, parents place limited value on the "academic quality" of the school as measured by things like student test scores, though parents with higher income tend to place a higher priority on test scores (Hastings et al. 2006). While these results suggest that choice plans may not lead to higher performance among poor children, there is some evidence that providing poor families with more transparent information about school quality does lead parents to select "better" schools in terms of academic performance (Hastings et al. 2007).

Even if school choice does not foster competition that increases the productivity of schooling overall, the opportunity to choose may still have a positive effect on students. Choice might allow a student to attend a school with better teachers, more resources, or more studious peers. Whether this re-sorting of children across schools improves average outcomes for all children, or simply changes the distribution of academic achievement across children, depends on whether some children gain more from "better" peer settings than others. Of course, if the goal is to reduce inequity in schooling opportunities and outcomes, and school choice provides choices to disadvantaged students who would not have them otherwise, then choice may still have served a worthwhile purpose. Some argue that choice provides poor parents the opportunities that wealthier parents have always had.

The availability of choice may also improve student outcomes if it allows parents to find schools that are a better "match" for their individual child's needs, which in principle could

allow most students to benefit from choice.[18] Some would argue that if choice allows parents to select the type of schools they want, the system should be considered a success. While this may be true from the parents' perspective, if society values certain outcomes more than others, then free choice on the part of parents may not maximize social welfare.

There are hundreds of studies on school choice, ranging from studies of charter schools to private voucher plans. However, the vast majority of research in this area suffers from a critical limitation, which researchers refer to as selection bias. Students who utilize school choice (whether it is a public magnet school, a charter school or a private school) differ from students who do not, both in terms of easily observable ways such as race, income and achievement level, but also likely in more subtle ways such as personal motivation or family support. If researchers do not account for such differences when comparing, for example, charter schools with neighborhood public schools, then the results can be misleading.

The best studies of the impact of attending a public school of choice on student achievement overcome the problem of selection bias mentioned above by focusing on choice schools that are oversubscribed and use random lotteries to determine admission. By comparing the outcomes of students who win and lose lotteries, the researchers can estimate the causal impact of attending a public school of choice. Looking at elementary and high schools in Chicago, Cullen et al. (2006, forthcoming) find that winning a choice lottery has no effect on a wide range of academic outcomes.[19] In a study of school choice lotteries in the Charlotte-Mecklenberg school district, Hastings et al. (2006a) find that, on average, students do not benefit from winning one of the lotteries and attending their preferred school. However, when they

[18] For example, the back-to-basics, discipline-oriented "academy" may be a good fit for some students, while other children may thrive more in an environment that provides more flexibility and autonomy. The match may reflect more mundane needs, including proximity to the parents' work or the availability of a particular elective.
[19] For high school students, the authors find some evidence that winning a choice lottery improves certain non-academic outcomes such as the likelihood of getting into trouble with the police.

examine children whose parents seemed to place a greater value on academics (based on the type of schools they chose), the authors find that these students do experience higher test scores from winning the public school choice lottery. Finally, as noted above, there is some evidence that career academies, one prominent type of public choice school, have positive effects on student outcomes (Cullen et al. 2005, Kemple and Willner, 2008).

A parallel literature analyzes the effect of attending a charter school on academic outcomes, and yields similarly mixed outcomes. Several studies find small negative or zero effects (Hanushek et al. 2005, Bifulco and Ladd 2006, Sass 2006). A recent random evaluation of New York City charters uses a lottery-based research design, comparing students who win and lose lotteries to oversubscribed charter schools, and finds robust positive effects (Hoxby and Murarka 2007). This approach removes concerns about selection bias, but has the drawback of focusing on a small number of particularly successful (or at least popular) charter schools.

Perhaps the most contentious form of school choice involves vouchers that allow students to attend private schools. Voucher experiments in Milwaukee, New York, Dayton and Washington, DC have been evaluated. These studies typically focus on comparing outcomes between students who are versus are not offered the chance to attend private schools. While the results in most cities were discouraging, the evaluation in NYC pointed to some modest test score gains for African-American children, although the magnitude of these results has been debated (Howell et al. 2002, Krueger and Zei 2004).

The most difficult area of school choice to analyze empirically is the claim that choice will foster competition that will, in turn, improve the productivity of all schools. Hoxby (2000) compared geographic areas, and found that those areas with more school choice for largely historical reasons (e.g. the presence of streams and rivers which made transportation difficult in earlier

times, and thus encouraged areas to divide themselves along the boundary of such waterways), had more productive schools, suggesting that greater choice is associated with greater efficiency in schooling. However, others have argued that these results are not particularly robust (Rothstein 2006, Hoxby 2006). Urquiola (2005) finds evidence of sorting: the "best" public school students left for the private sector. Areas with greater district choice tend to have a smaller fraction of students attending private schools, but these districts also have schools that are more racially homogenous.

In summary, there is mixed evidence on whether the opportunity to attend a choice school – public magnet, charter or private – has substantial academic benefits for poor children as well as on the question of whether large scale choice programs might improve the productivity of schools in a general. It is premature to make strong claims about whether such policies can have substantial benefits for disadvantaged children. Perhaps the main risk associated with expanded choice opportunities is the possibility of exacerbating the segregation of poor, minority or low-performing students within a subset of schools. This seems to have happened to some degree in New Zealand and Chile under the large-scale choice plans adopted there (Fiske and Ladd, 2000, Ladd and Fiske, 2001, Hsieh and Urquiola, 2006). However as both Neal (2002) and Ladd (2002) point out, the effects of any choice plan is likely to depend crucially on the details of key design questions, such as whether schools are allowed to select the best students from their applicant pools. Many of the current public school choice plans require over-subscribed schools to admit students on the basis of lotteries, providing some guard against this type of sorting. Additional experimentation by states and districts with different forms of public school choice (e.g., magnet or charter schools) would provide additional opportunities to learn more about their potential impacts.

V. CONCLUSIONS

The release of the landmark Coleman Report in 1966 fostered pessimism about the ability of schools to improve the life chances of poor children. The report found that a variety of school "inputs" such as teacher educational attainment and per pupil spending were only weakly correlated with student test scores, and raised questions about what schools could accomplish without broader changes in social policy. The Coleman Report and subsequent research pushed policymakers to consider outcome-based measures of success and spurred interest in reform strategies that focus on changing the incentives within the public school system.

A careful review of the empirical evidence, however, suggests a variety of policies that are likely to substantially improve the academic performance of poor children. Importantly, we found examples of successful programs or policies within each of the three broad categories outlined in the introduction. Targeted investment of additional resources in early childhood education, smaller class sizes and bonuses for teachers in hard-to-staff schools and subjects seem likely to pass a cost-benefit test, even without a fundamental re-organization of the existing public school system. At the same time, researchers have identified some ways of changing standard operating procedures within schools that can improve the outcomes of poor children even without large amounts of additional spending. Examples include the adoption of comprehensive school reform models like Success for All and Career Academies and the expansion of alternative certification opportunities. Finally, policies that seek to change incentives within schools, such as the accountability policies adopted by districts and states since the mid-1990s, offer some promise of improving schooling for poor children.

Given limited financial resources and perhaps even more limited political attention, it is unlikely that policymakers could adopt all of the "successful" practices discussed in this chapter.

Based on our read of the empirical literature, the following education policies seem likely to generate the largest gains in academic achievement among poor children:

1. Increased investments in early childhood education for poor children. Even though short-term gains in IQ or achievement test scores diminish over time, there is evidence of long-term improvement in a variety of outcomes such as educational attainment that will help children escape from poverty as adults. Increased investment in early childhood education may be particularly productive and important given the limited investment our society currently makes in the cognitive development of very young children.
2. Taking advantage of the opportunity provided by No Child Left Behind (NCLB) to better utilize accountability reforms to improve outcomes for poor children. NCLB was enacted in 2001 with bipartisan support, although it has received considerable criticism in recent years. But the debate over the existence of NCLB seems to miss a fundamental lesson we have learned about accountability in the past decade: the specific design of the program matters enormously. Potentially productive changes to NCLB (as well as to co-existing state or district accountability systems) include: adopting common achievement standards across states, focusing accountability on student growth rather than proficiency levels, providing states and districts with the flexibility to focus limited resources on the most needy schools, and reconciling federal and state accountability systems.
3. Provide educators incentives to adopt practices with a compelling research base while expanding efforts (including resources) to develop and identify effective instructional regimes. One of the lessons from the accountability movement is that highly

disadvantaged schools (and districts) often lack the capacity to change themselves. There is good evidence that certain comprehensive school reforms such as Success for All can raise achievement levels among poor children with little additional expenditures. Similarly, there is compelling evidence that alternative certification routes into teaching can expand the supply of teachers willing to work in hard-to-staff schools and subjects while maintaining, or perhaps even increasing, teacher quality. State and district officials could ensure that disadvantaged schools, particularly those that have continued to fail under recent accountability system, adopt instructional practices and related policies with a strong research base. This would help avoid the constant reinvention of the wheel. At the same time, the Federal government could help spur such advantages through more focused R&D spending (e.g., perhaps through "design competitions" as suggested by Slavin 1997), and governments at all levels could help increase the supply of high-quality practices by requiring schools to use programs that have been rigorously evaluated (Jacob and Ludwig, 2005).

4. Continue to support and evaluate a variety of public school choice options. The other noteworthy trend in output-based reform aside from NCLB is the spread of public school choice at the local level. While we believe that the current evidence on the benefits of public school choice is limited, the risk associated with these policies seems small so long as they are implemented in ways that do not substantially exacerbate school segregation along race or class lines. Hence, there could be considerable benefit to expanding magnet and charter schools, and to carefully evaluating the impact of these schools on the students they serve as well as the surrounding schools.

Most anti-poverty policies focus on lifting adults out of poverty. These policies are often controversial because of an unavoidable tension between the desire to help people who have been unlucky and the motivation to encourage hard work and punish socially unproductive behavior. In contrast, successful education policies can not only help reduce poverty over the long term by making poor children more productive during adulthood, but also foster economic growth that expands the "pie" for everyone. Educational interventions arguably also benefit from a compelling moral justification as well. Many people believe that disadvantaged children should not be punished for the circumstances into which they are born, and improved education policy is one of the best ways to prevent this from happening.

References

Aronson, Daniel, Barrow, Lisa and William Sander (2007). "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics* 25(1): 95-136.

Achilles, CM, Barbara A. Nye, Jayne B. Zaharias, and B. Dewayne Fulton (1993) "The lasting benefits study (LBS) in grades 4 and 5 (1990-91): A legacy from Tennessee's four-year (K-3) class size study (1985-1989), Project STAR." HEROS, Inc.

Ballou, Dale (1996) "Do Public Schools Hire the Best Applicants?" *The Quarterly Journal of Economics* 111(1): 97-133.

Ballou, Dale and Michael Podgursky (1995) "Recruiting Smarter Teachers." *The Journal of Human Resources* 30(2): 326-338.

Balter, D & Duncombe, W. (2005b). Teacher hiring practices in New York State school districts." Report prepared for the Education Finance Research Consortium, January.

Barnett, W. Steven, Cynthia Lamy and Kwanghee Jung (2005) "The effects of state pre-kindergarten programs on young children's school readiness in five states." Rutgers University, National Institute for Early Education Research.

Barnett, W. Steven and Leonard Masse (2007) "Comparative Benefit-Cost Analysis of the Abecedarian Program and Its Policy Implications." *Economics of Education Review*. 26: 113-125.

Belfield, Clieve R., Milagros Nores, Steve Barnett, and Lawrence Schweinhart (2006) "The High/Scope Perry Preschool Program: Cost-benefit analysis using data from the Age-40 Followup." *Journal of Human Resources*. 41(1): 162-190.

Besharov, Douglas J. (2002, February 27). Testimony before the Subcommittee on 21st Century Competitiveness of the Committee on Education and the Workforce. (www.welfareacademy.org/pubs/testimony-022702.pdf).

Bifulco, R. and Ladd, H. (2006). "The Impacts of Charter Schools on Student Achievement: Evidence from North Carolina," American Education Finance Association.

Bishop, John H. (1998) "Do Curriculum-Based External Exit Exam Systems Enhance Student Achievement?" Consortium for Policy Research in Education, University of Pennsylvania Graduate School of Education, Philadelphia

Bloom, Howard S. (1984) "Accounting for no-shows in experimental evaluation designs." *Evaluation Review* 8(2): 225-46.

Bloom, Howard S., Sandra Ham, Laura Melton and Julieanne O'Brien (2001) "Evaluating the Accelerated Schools Approach: A Look at Early Implementation and Impacts on Student

Achievement in Eight Elementary Schools." New York: MDRC.

Borhnstedt, George W. and Stecher, Brian M. (2002). *What Have We Learned about Class Size Reduction in California.* Capstone Report by CSR Research Consortium

Borman, Geoffrey D., & Hewes, Gina M. (2003) "The long-term effects and cost-effectiveness of Success for All." *Educational Evaluation and Policy Analysis* 24: 243-266

Borman, Geoffrey D., Robert E. Slavin, Alan C.K. Cheun, Anne M. Chamberlain, Nancy A. Madden, and Bette Chambers (2007) "Final reading outcomes of the National Randomized Field Trial of Success for All." *American Educational Research Journal*. 44(3): 701-731.

Boyd,Donald Pamela Grossman, Hamilton Lankford, Susanna Loeb, James Wyckoff (2005a). "How Changes in Entry Requirements Alter the Teacher Workforce and Affect Student Achievement." NBER Working Paper No. 11844

Boyd, Donald, Hamilton Lankford, Susanna Loeb, and James Wyckoff (2005b) "Explaining the Short Careers of High-Achieving Teachers in Schools with Low-Performing Students." *American Economic Review* 95(2): 166-172.

Campbell, Frances A. et al (2002) "Early Childhood Education: Young Adult Outcomes from the Abecedarian Project." *Applied Developmental Science* 6(1): 42-57.

Card, David and Krueger, Alan (1998) "School Resources and Student Outcomes," *The ANNALS of the American Acadamy of Political and Social Science* 559(39): 39-53.

Card, David and A. Abigail Payne (2002) "School finance reform, the distribution of school spending, and the distribution of student test scores." *Journal of Public Economics* 83: 49-82.

Card, David and Jesse Rothstein (forthcoming) "Racial Segregation and the Black-White Test Score Gap." *Journal of Public Economics.*

Carniero, Pedro and James J. Heckman (2003) "Human Capital Policy," In *Inequality in America: What Role for Human Capital Policies?* James J. Heckman and Alan B. Krueger. Cambridge, MA: MIT Press.

Carnoy, Martin and Susanna Loeb (2003) "Does External Accountability Affect Student Outcomes? A Cross-State Analysis." *Education Evaluation and Policy Analysis* 24(4). 305-331

Chiang, Hanley (2007) "How Accountability Pressure on Failing Schools Affects Student Achievement." Unpublished manuscript, Harvard University.

Clotfelter, Charles T., Elizabeth Glennie, Helen Ladd, and Jacob Vigdor (2006) "Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina." Cambridge, MA: National Bureau of Economic Research Working Paper 12285.

Clotfelter, Charles T., Ladd, Helen F. (1996) “Recognizing and Rewarding Success in Public Schools.” In Ladd, H., (Ed.), *Holding Schools Accountable: Performance-Based Reform in Education.* The Brookings Institution, Washington, D.C.

Clotfelter, Charles T., Helen F. Ladd, Jacob L. Vigdor (2006) "Teacher-Student Matching and the Assessment of Teacher Effectiveness." *Journal of Human Resources* 41: 778-820.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor (2007) “Who Teaches Whom? Race and the Distribution of Novice Teachers,” *Economics of Education Review* 24(4): 377-392.

CRSQ (2006). “Comprehensive School Reform Quality (CRSQ) Center Report on Elementary School Comprehensive School Reform Models.” American Institutes for Research.

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

Cook, Thomas D., Farah-Naaz Habib, Meredith Phillips, Richard A. Settersten, Shobha C. Shagle and Serdar M. Degirmencioglu (1999) “Comer's School Development Program in Prince George's County: A theory-based evaluation.” *American Educational Research Journal* 36(3): 543-597.

Cook, Thomas D., H. David Hunt, and Robert F. Murphy (2000) “Comer's School Development Program in Chicago: A theory-based evaluation.” *American Educational Research Journal* 37(2): 535-597.

Cooley, Jane (2007). “Desegregation and the Achievement Gap: Do Diverse Peers Help?” (under review).

Cullen, Julie and Randall Reback (2002) “Tinkering Toward Accolades: School Gaming under a Performance Accountability System.” Working paper, University of Michigan.

Cullen, Julie, Brian A. Jacob and Steven D. Levitt (2006) “The effect of school choice on student outcomes: Evidence from randomized lotteries.” *Econometrica*. 74(5): 1191-1230.

Currie, Janet (2001) “Early Childhood Education Programs.” *Journal of Economic Perspectives*. 15(2): 213-238.

Currie, Janet and Duncan Thomas (1995) “Does Head Start Make a Difference?” *American Economic Review*. 85(3): 341-364.

Deere, Donald and Wayne Strayer (2001) “Putting Schools to the Test: School Accountability, Incentives and Behavior.” Working paper, Department of Economics, Texas A&M University.

Deming, David (2007) “Early Childhood Intervention and Life-Cycle Skill Development.” Working Paper, Harvard University.

Dolton, Peter and Wilbert van der Klaauw (1999) "The Turnover of Teachers: A Competing Risks Explanation." *The Review of Economics and Statistics* 81(3): 543-550.

Downes, Thomas A. and David N. Figlio (1998) "School Finance Reforms, Tax Limits and Student Performance: Do Reforms Level-Up or Dumb Down?" Discussion Paper 98-05, Department of Economics, Tufts University.

Dynarski, Mark, Roberto Agodini, Sheila Heaviside, Timothy Novak, Nancy Care, Larissa Campuzano, Barbara Means, Robert Murphy, William Penuel, Hal Javitz, Deborah Emery, Willow Sussex (2007). "Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort." Mathematica Policy Research. For U.S. Department of Education, NCEE 2007-4005. http://www.mathematica-mpr.com/publications/PDFs/effectread.pdf

Figlio, David N. (2006) "Testing, crime and punishment." *Journal of Public Economics* 90(4-5): 837-851.

Figlio, David, Getzler, Lawrence (2002) "Accountability, Ability and Disability: Gaming the System?" Working paper, University of Florida.

Figlio, David N. and Cecilia Elena Rouse (2006) "Do Accountability and Voucher Threats Improve Low-Achieving Schools?" *Journal of Public Economics* 90 239-255.

Figlio, David , Winicki, Joshua, (forthcoming) "Food for Thought? The Effects of School Accountability on School Nutrition." *Journal of Public Economics*.

Fiske, Edward B. and Helen F.Ladd (2000) *When Schools Compete: A Cautionary Tale*. Washington, DC: Brookings Institution Press.

Frederisksen, Norman (1994) "The Influence of Minimum Competency Tests on Teaching and Learning." Educational Testing Service, Policy Information Center, Princeton NJ.

Garces, Eliana, Duncan Thomas, and Janet Currie (2002) "Longer Term Effects of Head Start." *American Economic Review*. 92(4): 999-1012.

Glazer, Nathan (1986) "Education and training programs and poverty." In Sheldon Danziger and Daniel Weinberg, Eds. *Fighting Poverty: What Works and What Doesn't*. Cambridge, MA: Harvard University Press. Pp. 152-172.

Glazerman, Steven, Tom Silva, Nii Addy, Sarah Avellar, Jeffrey Max, Allison McKie, Brenda Natzke, Michael Puma, Patrick Wolf, & Rachel Ungerer Greszler, (2006) "Options for studying teacher pay reform using natural experiments." Washington, DC: Mathematica Policy Research, Inc. No. ED-04-CO-0112/0002.

Glazerman, Steven, Daniel Mayer, and Paul Decker (2006) "Alternative Routes to Teaching: The Impacts of Teach for America on Student Achievement and Other Outcomes." *Journal of Policy Analysis and Management*. 25(1): 75-96.

Gordon, Nora E. (2004) "Do Federal funds boost school spending? Evidence from Title I." *Journal of Public Economics* 88(9-10): 1771-1792.

Gordon, Robert, Thomas J. Kane, and Douglas O. Staiger (2006) "Identifying effective teachers using performance on the job." Washington, DC: Brookings Institution, Hamilton Project Discussion Paper 2006-01.

Gormley, William T. and Ted Gayer (2005) "Promoting School Readiness in Oklahoma: An Evaluation of Tulsa's Pre-K Program." *Journal of Human Resources*. 40(3): 533-558.

Gormley, William T., Ted Gayer, Deborah Phillips and Brittany Dawson (2005) "The effects of universal pre-K on cognitive development." Working Paper, Georgetown University, Center for Research on Children in the United States.

Grissmer, David and Ann Flanagan, (1998) "Exploring Rapid Achievement Gains in North Carolina and Texas." National Education Goals Panel, Washington, D.C.

Grissmer, David, et al (2000) "Improving Student Achievement: What NAEP Test Scores Tell Us." RAND Corporation, Santa Monica, CA MR-924-EDU.

Guarino, Cassandra M., Lucrecia Santibanez, and Glenn A. Daley, (2006) "Teacher Recruitment and Retention: A Review of the Recent Empirical Literature." *Review of Educational Research* 76(2): 173-208

Guryan, Jonathan (2004) "Desegregation and Black Dropout Rates," *American Economic Review*, 94(4): 919-943.

Haney, Walt (2000) "The Myth of the Texas Miracle in Education." *Education Policy Analysis Archives* 8(41).

Hanushek, Eric A., John F. Kain and Steven G. Rivkin (2004). "Why Public Schools Lose Teachers" *Journal of Human Resources* 39(2): 326-354.

Hanushek, Eric, Kain, John F., O'Brien, Daniel M. and Steven G. Rivkin (2005). "The Market for Teachers." NBER Working Paper 11154.

Hanushek, Eric, Kain, John, Rivkin, Steven, and Branch, G. (2005). "Charter School Quality and Parental Decision Making with School Choice," NBER Working Paper No. 11252.

Hanushek, Eric A. and Margaret E. Raymond (2005) "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis and Management* 24(2): 297-327.

Hanushek, Eric, and Rivkin, Steven (2007). "Pay, Working Conditions, and Teacher Quality," *The Future of Children* 17(1): 69-86.

Harris, Douglas N. and Tim R. Sass. (2007). "What Makes for a Good Teacher and Who Can Tell?" Unpublished manuscript, Florida State University, July

Harris, Douglas N., Rutledge, S., Ingle, W.K., Thompson, C. (2006). "When Supply Meets Demand: Principal Preferences and Teacher Hiring." Paper prepared for the 2006 annual meeting of the American Educational Research Association, San Francisco, CA.

Hastings, Justine S., Thomas Kane, and Douglas Staiger (2006) "Preferences and heterogeneous treatment effects in a public school choice lottery." Cambridge, MA: NBER Working Paper 12145.

Hastings, Justine S., Thomas J. Kane, and Douglas O. Staiger (2006). "Parental Preferences and School Competition: Evidence from a Public School Choice Program." NBER Working Paper No. 11805

Hastings, Justine S., Richard Van Weelden, and Jeffrey Weinstein (2007). "Preferences, Information, and Parental Choice Behavior in Public School Choice." NBER Working Paper No. 12995.

Herlihy, Corinne and Janet Quint (2006). "Emerging Evidence on Improving High School Student Achievement and Graduation Rates: The Effects of Four Popular Improvement Programs." National High School Center Research Brief. http://www.betterhighschools.org/docs/NHSC_EmergingEvidenceBrief_111606Final.pdf

Hill, Heather C. (2007). "Learning in the Teaching Workforce." *The Future of Children*, 17(1) (Special Issue: Excellence in the Classroom.): 111-127.

Howell, William, Paul E. Peterson with Patrick J. Wolf (2002). "The Education Gap: Vouchers and Urban Schools." Brookings Institution Press.

Hoxby, Caroline (2000). "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics* 115(4): 1239-1285.

Hoxby, Caroline M. and Sonali Murarka (2007). "Charter Schools in New York City: Who Enrolls and How they Affect Their Students' Achievement." Working Paper.

Hsieh, Chang-Tai and Miguel Urquiola (2006) "The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program." *Journal of Public Economics*. 90: 1477-1503.

Imazeki, Jennifer (2007). "Attracting and Retaining Teachers in High-Needs Schools: Do Financial Incentives Make Financial Sense?" Working Paper. San Deigo State University, Department of Economics

Jacob, Brian (2001) "Getting Tough? The Impact of Mandatory High School Graduation Exams on Student Outcomes." *Educational Evaluation and Policy Analysis* 23(2): 99-122.

Jacob, Brian (2005) "Accountability, Incentives and Behavior: the Impact of High-Stakes Testing in the Chicago Public Schools." *Journal of Public Economics* 89(5-6): 761-796.

Jacob, Brian (2007). "The Challenges of Staffing Urban Schools with Effective Teachers." *The Future of Children* 17(2). (Special Issue: Excellence in the Classroom. Policies to Improve the Teacher Workforce.)

Jacob, Brian A. (2008). "The Effect of Employment Protection on Productivity: Evidence from a Change in Teacher Dismissal Policy in the Chicago Public Schools." Working paper.

Jacob, Brian, Levitt, Steven (2003) "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics* 117(3): 843-878.

Jacob, Brian and Lars Lefgren (2007) "What Do Parents Value in Education? An Empirical Investigation of Parents' Revealed Preferences for Teachers." *Quarterly Journal of Economics* 122(4):1603-1637.

Jacob, Brian, Thomas Kane, Jonah Rockoff and Douglas Staiger (2008). "What Pre-Employment Characteristics Predict Teacher Effectiveness: Evidence from New York City." Working Paper.

Jargowsky, Paul A. (1996) "Take the Money and Run: Economic Segregation in U.S. Metropolitan Areas." *American Sociological Review* 61(6): 984-998.

Jargowsky, Paul A. (2003) "Stunning Progress, Hidden Problems: The Dramatic Decline of Concentrated Poverty in the 1990s." Report for the Living Cities Census Series. Washington, DC: The Brookings Institution.

Jencks, Christopher (1986) "Education and training programs and poverty." In Sheldon Danziger and Daniel Weinberg, Eds. *Fighting Poverty: What Works and What Doesn't.* Cambridge, MA: Harvard University Press. Pp. 173-179.

Jepsen, Christopher and Steven G. Rivkin (2002) "What is the Tradeoff Between Smaller Classes and Teacher Quality?" NBER Working Paper 9205.

Kahlenberg, Richard D. (2001) "Review of Brown v. Board of Education, by James T. Patterson." *The American Prospect*. May 20, 2001.

Kane, Thomas J. and Douglas Staiger (2002a) "The promise and pitfalls of using imprecise school accountability measures." *Journal of Economic Perspectives*. 16(4): 91-114.

Kane, Thomas J. and Douglas Staiger (2002b) "Volatility in school test scores: Implications for

test based accountability systems." In Diane Ravitch, Editor. *Brookings Papers on Education Policy*. Washington, DC: Brookings Institution Press.

Kane, Thomas J., Jonah Rockoff and Douglas Staiger (2006) "What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City" NBER Working Paper No. 12155, (Submitted for publication, under review)

Kemple, James J. and Judith Scott-Clayton (2004) *Career Academies: Impacts on Labor Market Outcomes and Educational Attainment*. New York: MDRC

Kemple, James J., Corinne M. Herlihy, and Thomas J. Smith. (2005) *Making Progress Toward Graduation: Evidence From the Talent Development High School Model*. New York: MDRC

Kemple, James J. and Cynthia J. Willner (2008) *Career Academies: Long-Term Impacts on Labor Market Outcomes, Educational Attainment, and Transitions to Adulthood*. New York: MDRC.

Klein, Stephen, Laura Hamilton, et al (2000) "What Do Test Scores in Texas Tell Us?" RAND Corporation, Santa Monica, CA.

Kling, Jeffrey R., Jens Ludwig and Lawrence F. Katz (2005) "Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment." *Quarterly Journal of Economics*, 120(1): 87-130.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz (2007) "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1): 83-119.

Knudsen, Eric I., James J. Heckman, Judy L. Cameron, and Jack P. Shonkoff (2006) "Economic, neurobiological, and behavioral perspectives on building America's future workforce." *Proceedings of the National Academy of Sciences*. 103: 10155-10162.

Koretz, Daniel, Linn, Robert, et al (1991) "The Effects of High-Stakes Testing: Preliminary Evidence About Generalization Across Tests." American Educational Research Association, Chicago.

Koretz, Daniel., Barron, Sheila (1998) "The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS)." RAND Corporation, Santa Monica, CA.

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

Krueger, Alan B. (2003) "Economic considerations and class size." *Economic Journal* 113: 3-33.

Krueger, Alan B. and Diane M. Whitmore (2001). "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project

STAR." *Economic Journal* 111: 1-28.

Krueger, Alan B. and Diane M. Whitmore (2002). "Would Smaller Classes Help Close the Black-White Achievement Gap?" In *Bridging the Achievement Gap*, John Chub and Tom Loveless (eds.), Washington, DC: Brookings Institute Press.

Krueger; Alan B and Pei Zhu (2004). "Another Look at the New York City School Voucher Experiment." *The American Behavioral Scientist*; 47(5): 658

Ladd, Helen F. (1996). *Holding Schools Accountable: Performance-Based Reform in Education*. Washington, D.C.: The Brookings Institution. (Chapter 1 & 2, pages 1-19, 23-63)

Ladd, Helen F. (2002) "School vouchers: A critical view." *Journal of Economic Perspectives*. 16(4): 3-24.

Ladd, Helen F., Rosemary Chalk, and Janet S. Hansen, editors. Equity and Adequacy in Education Finance: Issues and Perspectives. National Academy Press, Washington, D.C., 1999.

Ladd, Helen F. and Edward B. Fiske (2001) "The uneven playing field of school choice: Evidence from New Zealand." *Journal of Policy Analysis and Management*. 20(1): 43-63.

Lankford, Hamilton, Susanna Loeb, and James Wyckoff (2002). "Teacher Sorting and the Plight of Urban Schools," *Educational Evaluation and Policy Analysis* 24: 37–62.

Lavy, Victor (2007). "Using Performance-Based Pay to Improve the Quality of Teachers", The Future of Children, Spring 2007, 87-110.

Lazear, Edward P. (2001) "Educational Production." *The Quarterly Journal of Economics* 116(3): 777-803.

Levin, Henry (1977) "A decade of policy developments in improving education and training for low-income populations." In Robert Haveman, Ed. *A Decade of Federal Anti-Poverty Policy: Achievements, Failures, and Lessons*. New York: Academic Press.

Levin, Henry (2002). "Issues in Designing Cost-Effectiveness Comparisons of Whole-School Reforms," In Levin and McEwan, *Cost-Effectiveness and Educational Policy*, AEFA Yearbook 2002 (Eye on Education 2002), pp. 89-92.

Levin, Jessica and Meredith Quinn (2003). Missed Opportunities: How We Keep High-Quality Teachers Out of Urban Classrooms. New York, NY: The New Teacher Project, www.tntp.org

Levin, Jessica, Jennifer Mulhern, and Joan Schunck, (2005) "Unintended Consequences: The Case for Reforming the Staffing Rules in Urban Teachers Union Contracts." New York, NY: The New Teacher Project: www.tntp.org.

Linn, Robert, M. Elizabeth Graue, et al (1990) "Comparing State and District Results to National

Norms: The Validity of the Claim that 'Everyone is Above Average." *Educational Measurement: Issues and Practice* 9(3): 5-14.

Levin, Henry (2002) "Issues in Designing Cost-Effectiveness Comparisons of Whole-School Reforms," in Levin and McEwan, COST-EFFECTIVENESS AND EDUCATIONAL POLICY, AEFA Yearbook. *Eye on Education* 89-92.

Lewin, Tamar (2007, June 29). Across U.S., a New Look at School Integration Efforts . *The New York Times*.

Loeb, Susanna and Marianne E. Page (2000) "Estimating the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation." *The Review of Economics and Statistics*. 82(3): 393-408.

Love, John M., Ellen Eliason Kisker, Christine M. Ross, Peter Z. Schochet, Jeanne Brooks-Gunn, Diane Paulsell, Kimberly Boller, Jill Constantine, Cheri Vogel, Allison Sidle Fuligni, and Christy Brady-Smith. (2002) "Making a Difference In The Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start." Mathematica Policy Research Report.

Ludwig, Jens and Douglas L. Miller (2007) "Does Head Start Improve Children's Life Chances? Evidence from a Regression-Discontinuity Design." *Quarterly Journal of Economics*. 122(1): 159-208.

Ludwig, Jens and Deborah A. Phillips (2007) "The Benefits and Costs of Head Start." NBER Working Paper 12973.

Lutz, Byron F. (2005) "Post Brown vs. the Board of Education: The Effects of the End of Court-Ordered Desegregation." Finance and Economics Discussion Series 2005-64. Washington: Board of Governors of the Federal Reserve System.

Magnuson, Katherine A., Christopher J. Ruhm, and Jane Waldfogel (2004) "Does prekindergarten improve school preparation and performance?" Cambridge, MA: National Bureau of Economic Research Working Paper 10452.

Magnuson, Katherine and Elizabeth Votruba-Drzal (2009) "Enduring influences of childhood poverty." *Changing Poverty*, edited by Maria Cancian and Sheldon Danziger. NY: Russell Sage Foundation.

Murray, Sheila E., William N. Evans, and Robert M. Schwab (1998) "Education-finance reform and the distribution of education resources." *The American Economic Review* 88(4): 789-812.

National Center for Education Statistics (2007). "Trends in the Use of School Choice: 1993 to 2003." NCES 2007-045.

National Center for Education Statistics (2007) "The Nation's Report Card, Mathematics 2007: National Assessment of Educational Progress at Grades 4 and 8." NCES 2007-494. U.S. Department of Education, Institute of Education Sciences.

National Center for Education Statistics (2007) "The Nation's Report Card, Reading 2007: National Assessment of Educational Progress at Grades 4 and 8." NCES 2007-496. U.S. Department of Education, Institute of Education Sciences.

Neal, Derek A. (2002) "How vouchers could change the market for education." *Journal of Economic Perspectives*. 16(4): 25-44.

Neal, Derek A. and Diane Schanzenbach (2007). "Left Behind by Design: Proficiency Counts and Test-Based Accountability." NBER Working Paper 13293.

Nye, B., Fulton B.D., J. Boyd-Zaharias, and V.A. Cain (1995) "The lasting benefits study, eighth grade technical report." Nashville, TN: Center for Excellence for Research in Basic Skills, Tennessee State University.

Nye, B., Larry V. Hedges, and S. Konstantopoulos (2002) "Do low achieving students benefit more from small classes? Evidence from the Tennessee class size experiment." *Educational Evaluation and Policy Analysis*. 24: 201-217.

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

Podgursky, M. and Springer, M.G. (2007). "Teacher Performance Pay: A Review." Journal of Policy Analysis and Management, 26(4).

Puma, Michael, Stephen Bell, Ronna Cook, Camilla Heid, Michael Lopez, et al. (2005) "Head Start Impact Study: First Year Findings." Westat: Report Prepared for the U.S. Department of Health and Human Services.

Quint, Janet, Howard Bloom, Alison Rebeck Black, LaFleur Stephens, and Teresa Akey. (2005) "First Things First: The Challenge of Scaling Up Educational Reform." New York: MDRC.

Ramey, Craig T. and Frances A. Campbell, (1979) "Compensatory Education for Disadvantaged Children," *The School Review* 87(2): 171-189.

Ramey, Craig T. and Frances A. Campbell (1984) "Preventive Education for High-Risk Children: Cognitive Consequences of the Carolina Abecedarian Project." *American Journal of Mental Deficiency* 88(5): 515-23.

Richards, Craig E. and Tian Ming Sheu (1992). "The South Carolina School Incentive Reward Program: A Policy Analysis." *Economics of Education Review* 11(1): 71-86.

Rivkin, Steven G. (2000). "School Desegregation, Academic Attainment, and Earnings." *The Journal of Human Resources*, 35(2): 333-346.

Rothstein, Jesse (2007) “Do value-added models add value? Tracking, fixed effects and causal inference.” Working Paper, Princeton University Department of Economics.

Rouse, Cecilia E., Hannaway, Jane, Goldhaber, Dan D. and Figlio, David N. (2007). "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure". NBER Working Paper No. 13681.

Rumberger, Russel W. and Gregory J. Palardy (2005). “Does Segregation Still Matter? The Impact of Student Composition on Academic Achievement in High School.” *The Teachers College Record* 107(9): 1999-2045.

Sanbonmatsu, Lisa, Jeffrey R. Kling, Greg J. Duncan, and Jeanne Brooks-Gunn (2006) “Neighborhoods and academic achievement: Results from the MTO Experiment.” *Journal of Human Resources*. 41(4): 649-691.

Sass, Tim (2006). Charter Schools and Student Achievement in Florida. *Education Finance and Policy*. 1: 91-122

Schanzenbach, Diane Whitmore (2006/7) “What have researchers learned from Project STAR?” *Brookings Papers on Education Policy*. Edited by Thomas Loveless and Frederick Hess. Washington, DC: Brookings Institution Press. Pp. 205-228.

Schweinhart, Lawrence J., Jeanne Montie, Zongping Xiang, W. Steven Barnett, Clive R. Belfield and Milagros Nores (2005) *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40*. Ypsilanti, Michigan: High/Scope Press.

Scafidi, Benjamin, David L. Sjoquist, and Todd R. Stinebrickner (2007). “Race, poverty and teacher mobility.” *Economics of Education Review* 26(2):145-159.

Shepard, Lorrie (1990) “Inflated Test Score Gains: Is the Problem Old Norms or Teaching the Test?” *Educational Measurement: Issues and Practice* 9(3): 15-22.

Shonkoff, Jack, and Deborah Phillips, eds. (2000) “From Neurons to Neighborhoods: The Science of Early Childhood Development.” Washington, D.C.: National Academy Press. Spence, A. Michael. 1974. *Market signaling: informational transfer in hiring and related screening processes*. Cambridge: Harvard University Press.

Slavin, Robert E. (1997) “Design competitions: A proposal for a new federal role in educational research and development.” *Educational Researcher*. 26(1): 22-28.

Smith, Stephen Samuel, Roslyn Arlin Mickelson (2000) “All that Glitters is Not Gold: School Reform in Charlotte-Mecklenburg.” *Educational Evaluation and Policy Analysis* 22(2): 101-127.

Snipes, Jason C., Glee Ivory Holton, Fred Doolittle, and Laura Sztejnberg (2006) “Striving for Success: The Effect of Project GRAD on High School Student Outcomes in Three Urban School Districts.” New York: MDRC.

Stecher, Brian M., Barron, Sheila I., (1999) "Quadrennial Milepost Accountability Testing in Kentucky." Center for the Study of Evaluation, University of California, Los Angeles.

Steele, Jennifer, Murnane, Richard, and Willet, John (2008). "Do Financial Incentives Help Low-Performing Schools Attract and Retain Academically Talented Teachers? Evidence from California." Working Paper.

Stinebrickner, Todd R. "A Dynamic Model of Teacher Labor Supply," *Journal of Labor Economics*, 19(1) January 2001, 196-230.

Stinebrickner, Todd R. (2002). "An Analysis of Occupational Change and Departure From the Labor Force - Evidence of the Reasons That Teachers Leave." *Journal of Human Resources,* 37 (1): 192-216.

Strauss, R. P., Bowes, L. R., Marks, M. S., & Plesko, M. R. (2000). Improving teacher preparation and selection: Lessons from the Pennsylvania experience. *Economics of Education Review*, 19(4), 387-415.

Toenjes, Lawrence A., A. Gary Dworkin, Jon Lorence, and Antwanette N. Hill (2000) "The Lone Star Gamble: High Stakes Testing, Accountability and Student Achievement in Texas and Houston." Mimeo, The Sociology of Education Research Group (SERG), Department of Sociology, University of Texas.

Urquiola, Miguel (2005). "Does School Choice Lead to Sorting? Evidence from Tiebout Variation" *American Economic Review* 95(4): 1310-1326.

Van der Klaauw, Wilbert (2008) "Breaking the link between poverty and low student achievement: An evaluation of Title I." *Journal of Econometrics* 142(2): 731-756.

Vigdor, Jacob and Thomas Nechyba (2006). "Peer Effects in North Carolina Schools" in School and the Equal Opportunity Problem, (eds. Ludger Woessmann, Paul E. Peterson). MIT Press.

Weiner, David A., Byron F. Lutz, and Jens Ludwig (2007). "The Effects of School Desegregation on Crime." Working Paper, University of Chicago.

What Works Clearinghouse (2007) "Intervention Report: Everyday Mathematics." U.S. Department of Education.

What Works Clearinghouse (2007) "Topic Report: Middle School Math." U.S. Department of Education.

Winfield, Linda F. (1990) "School Competency Testing Reforms and Student Achievement: Exploring a National Perspective." *Educational Evaluation and Policy Analysis* 12(2): 157-173.

Yinger, John (2004). *Helping Children Left Behind*. The MIT Press.