



Association for Public Policy Analysis and Management

**2009 Spencer Foundation Lecture
in Education Policy and Management**

“Educational Policy Research: Progress, Puzzles, and Challenges”

By Richard J. Murnane¹

© 2009 Association for Public Policy Analysis and Management.
All rights reserved.

¹ Thompson Professor of Education and Society, Harvard University Graduate School of Education. This paper is a revised version of the Spencer Foundation Lecture in Education Policy and Management presented by the author at the 2009 APPAM Fall Research Conference in Washington D.C. on November 5, 2009. The author is grateful to the Spencer Foundation for funding much of the research on which this lecture is based.

“Educational Policy Research: Progress, Puzzles, and Challenges”

By Richard J. Murnane

Introduction

I am honored to have been chosen to give the Spencer Lecture in Educational Policy and Management this year, especially to follow Rick Hanushek and Sunny Ladd, from whose work I have learned so much. My talk addresses three topics. The first concerns the reasons that American education, which served the country so well for the first three-quarters of the twentieth century, is not adequate today, and is especially not successful in providing disadvantaged children with the skills they need to escape poverty. The second concerns lessons from policy analyses about the efficacy of alternative approaches to improving American education, especially the education of its most disadvantaged children. I will make the case that research, much of it presented at APPAM meetings, provides encouraging evidence on several policy approaches. At the same time, the research reveals puzzles that I believe we need to take seriously in thinking about future research to inform the design of the next round of policies. Finally, I will suggest some directions for policy research that come from taking the puzzles seriously. In particular, I will argue the importance of learning more about the factors that facilitate and those that hinder the development of schools as organizations that continuously improve their performance.

American Education and Disadvantaged Children Today

As Claudia Goldin and Larry Katz describe so well in their 2008 book, *The Race Between Education and Technology*, the educational attainments of the American population grew very rapidly during the first three quarters of the 20th century. This fueled economic growth and dramatically reduced education-related earnings differentials. In their words, this was a period

of “growing together,” meaning that all segments of the population increased educational attainments and this enabled all segments to reap benefits from the nation’s economic growth. In contrast, Claudia and Larry characterize the decades since 1980 as a period of “growing apart,” a time in which income inequality grew rapidly, and education no longer provided the mechanism for all to benefit from the nation’s economic growth.

So, what is different about today’s world that has turned American education, which Claudia and Larry depict as a great strength of American society during most of the 20th century, into a serious problem today? While there are many contributing factors, I want to emphasize four.

The first are technological changes that have altered skill demands. In particular, advances in computer technology have dramatically reduced the demand for workers who can simply follow a set of well-defined instructions – the kinds of work that American education prepared several generations of students to do during the 20th century. These same technological changes have increased the demand for non-routine problem-solving skills and a variety of interactive skills that Frank Levy and I called expert thinking and complex communication (Levy and Murnane 2004). It is important to keep in mind that the U.S. schools never provided more than a modest minority of students with these skills. Thus, a big part of the educational challenge today is that American schools are trying to do something that the nation’s schools never did in the past. Moreover, the poverty into which a large percentage of urban children are born hinders the development of the language and communications skills that are so important in workplaces where computers are doing more and more of the routine tasks.

A second change is that other countries have dramatically increased the educational attainments of their young people. While 40 years ago, the U.S. was a leader in the educational

attainments of its population, this is no longer true (OECD 2009). This matters not only because the U.S. competes with other countries in product and service markets, but also because technology makes it possible for American firms to outsource work to other countries. The net effect is more competition for jobs, especially for workers who lack strong analytical skills and communication skills.

The third change is that the education sector must compete for talent in ways that it did not 40 years ago when job opportunities for female and minority college graduates were much more limited than they are today (Corcoran, Evans, and Schwab 2004). This is especially tough because the skills needed to teach all students to excel at expert thinking and complex communication are greater than those required to prepare students to work effectively in jobs that consisted primarily of following directions.

The fourth change is an increase in residential segregation that leaves urban school systems without the support of middle class parents, and court rulings that leave urban districts solely responsible for the education of high concentrations of economically disadvantaged students with great needs (Watson 2009). Work in most high poverty schools is tough and usually disheartening. The net effect is that American children most in need of the nation’s best teacher are the least likely to get them.

The growing importance of particular skill sets in determining labor market earnings makes access to a good education particularly important today. However, the same inequalities that make education so important as an engine of intergenerational mobility hinder access to a good education. To my mind, this contributes to the disturbing pattern that the rate of intergenerational economic mobility in the U.S. is lower today than it is in most other high income countries (Bjorklund and Jantti 2009).

So, to repeat the argument: changes in the American economy have increased the importance of a good education in determining labor market outcomes. However, increased inequality in many aspects of American life – including earnings and wealth – undermines the ability of families at the bottom of the income distribution to obtain a good education for their children.

Lessons from Recent Policy Alternatives

So what is to be done? Preschool interventions to improve children’s development during the early years of life are clearly important. So are initiatives to reduce the high concentrations of poverty and violence in which many families live. However, I am not going to say more about these important areas for public policy.

Instead, I am going to focus on improving K-12 education, especially for children living in poverty. I will discuss evidence, much of it presented at APPAM meetings, that bears on several types of policy initiatives, recognizing that they are complements rather than substitutes. In each case, there is some positive evidence. However, there are also puzzles and questions that need to be taken seriously if policies are to improve.

One set of initiatives aims to attract more academically talented college graduates to teaching. The most well known of these initiatives is *Teach for America*, which currently places approximately 4000 academically strong graduates in public schools serving high concentrations of economically disadvantaged children. The encouraging evidence from TFA is the large number of young, academically strong college graduates willing to do this important and difficult work.

The evidence that I personally find puzzling is that *Teach For America* participants are not much better, on average, if they are better at all, than teachers who enter the classroom from other routes. For example, Jonah Rockoff and his colleagues, using data from New York City, found no statistically significant difference between the average effectiveness of *TFA* participants and those who entered teaching through more conventional pathways (Rockoff et al. 2008). The random assignment study conducted by Mathematica that compared the effectiveness of *TFA* participants and other teachers in the same schools found that students taught by *TFA* participants had a little higher achievement in mathematics, but no higher reading achievement than students taught by the other teachers in these schools, who were among the least well prepared teachers in the nation (Decker, Mayer, and Glazerman 2004). I see this as a puzzle because I teach more than a dozen *TFA* graduates every year. They are remarkable people: smart, well educated, and committed. They learn rapidly, work hard, and ask probing questions. When policymakers talk about attracting greater talent to the teaching profession, it is hard for me to conceive of young people more able than those who join *TFA*. Yet the evidence does not support their relative effectiveness.

One potential explanation is that *TFA* participants are typically placed in schools with very weak teaching staffs. As a result, they obtain little guidance about how to do well the difficult work of teaching children born into poverty. A recently published paper by Kirabo Jackson and Elias Bruegmann provides support for this potential explanation (Jackson and Bruegmann 2009). The authors show that the success of individual teachers in improving students' skill levels depends on the effectiveness of their colleagues, and that this is especially true for beginning teachers. This raises the question of whether placing *TFA* participants and other academically talented novices in schools with skilled senior colleagues committed to

working together would produce significantly better education for disadvantaged children and greater growth in the skills of the academically talented novices.

A second set of policy initiatives aims to base pay and/or job security on estimates of teachers’ value-added, as measured by their students’ gains on standardized tests. The logic underlying these proposals is that teachers’ vary enormously in effectiveness and their performance is not predicted by the attributes rewarded in teacher salary schedules (Rivkin, Hanushek, and Kain 2005). Tom Kane’s and Doug Staiger’s 2008 paper showing that value-added models fitted with non-experimental data provide results that are not substantively different from those obtained from a random assignment experiment suggests that much can be learned from the results of well-specified value-added models.

Yet, a puzzling aspect of the evidence is the substantial year to year variation in estimates of value-added for the same teachers. A group of researchers at Rand estimates this to be about 50% of the total variation not due to noise, and that this is the case among experienced teachers as well as novices (McCaffrey et al. 2009). To my mind, this challenges in an important way the idea that a good school is simply a building housing all good teachers.

One potential explanation for the year-to-year variation in estimated performance is variation in class composition, especially the impact of one or more emotionally disturbed students who frequently disrupt instruction. Every teacher I have ever talked with has at least one story about the year “Jamie” or “Henry” – usually one or more boys – totally disrupted the class and prevented much learning from taking place.

A second potential explanation is that a teacher’s success in improving students’ scores in a particular year may depend on the extent to which her methods of teaching difficult skills like the multiplication of fractions are consistent with those used by her students’ previous classroom

teachers – or are different from the methods used by previous teachers – something that confuses struggling learners and hinders their achievement growth. In too many schools, the extent to which instruction is consistent across years is the luck of the draw since most teachers work as solo practitioners in isolation.

Would the year -to-year variation in teachers’ value-added be less in schools in which teachers worked together to provide consistency in instructional methods, and assumed joint responsibility for dealing with troubled students. Of course, the corollary to this question is more important: would student achievement growth be consistently higher in such schools?

You have now heard me suggest a particular interpretation of the puzzles in the research evidence on the effectiveness of individual teachers. This interpretation is that a necessary condition for educating disadvantaged children well is that schools be learning organizations in which talented adults work together to make instruction more consistent and who take joint responsibility for the development of all students. I make this suggestion even though I am aware that most quantitative studies find large teacher effects on student achievement but little or no school effects. I believe this pattern stems from the reality that most schools are not learning organizations. Instead, they are buildings in which individual practitioners work in substantial isolation.

To illustrate what I mean by a “learning organization,” I turn to a story about the treatment of cystic fibrosis that the surgeon, Atul Gawande, published in *The New Yorker* a few years ago (Gawande 2004). Gawande explains that there are 117 centers for the treatment of cystic fibrosis in the United States, and that the Cystic Fibrosis Foundation has kept records on their performance for several decades. There are two striking patterns in the evidence. First, the average performance of the centers, as measured by the median age to which cystic fibrosis

patients survive, has increased markedly in all centers. Gawande attributes this in large part to the rapid adoption of new treatments that were shown to be effective in randomized clinical trials.

The second pattern is remarkable variation among the centers in their performance, and the performance rankings are quite stable from year to year. Moreover, the explanation for the variation in performance among centers does not lie in differences in standardized treatments – all centers quickly adopt new treatments shown to be more effective than alternatives in clinical trials. This pattern led Gawande to spend some time at the center that consistently ranks first, which is located in Minneapolis. He writes about observing one of the doctors meeting with a 17-year-old female patient, Janelle, for her routine 3 month checkup. Prior to the meeting, Janelle had had a lung function test at the center. This is important because lung function is a critical indicator of the health of people with cystic fibrosis. The reason is that the disease leads to a thickening of mucus in the lungs that reduces lung function and ultimately leads to death. To prevent this, patients need to do twice-a-day 30 minute exercises to remove mucus from the lungs. Because the exercises are arduous and time-consuming, patients tend to skip them. They typically do not notice the decline in lung function for a significant period of time. However, by the time they do, a lot of damage has been done.

The results of Janelle’s lung function tests had been running at greater than 100 percent – better than the average of individuals without cystic fibrosis. This time, her lung function was at 90%. The physician asked Janelle if she had been doing her exercises and Janelle assured him that she had and that she felt fine. So, in terms of following prescribed procedures, the doctor had asked the right questions. However, he was not satisfied. He pressed on with more questions. Eventually, Janelle confessed that she was spending nights with a new boy friend and had not

been doing her exercises regularly. The doctor then negotiated a new treatment plan with Janelle that she said she could carry out. However, the doctor insisted that Janelle enter the hospital immediately and do intensive treatments until her lung function had completely recovered. She then needed to come for much more frequent check-ups to see whether the new treatment regime was working.

Gawande then turns from the vignette about Janelle to a description of what separates the Minneapolis center from other less successful treatment centers. He points to several things. There is a weekly meeting of all staff members in which they review everyone's care for all patients. The goal is that all staff members have the knowledge needed to care for all patients and feel the responsibility to do so. There is close attention to early warning signals of problems and intense interventions at the first sign of a problem. Also, the director insists of a high degree of uniformity of practice so that patients get the same attention to early signals of problems and the same probing questions no matter which clinician they see. This means that staff members need to learn from each other how to do these things, and so best practice is developed collectively and uniformly adhered to.

An emerging consensus among educators is that schools that are effective in educating disadvantaged children are learning organizations with properties similar to those of the Minneapolis cystic fibrosis treatment center. It is easy to see why. First, consistency of treatment is at least as important in educating children as it is in treating cystic fibrosis. It is thoroughly confusing to struggling students when teachers use different methods to teach the same skills, especially when their parents lack the resources to resolve the confusion. Of course, to obtain consistent instruction, teachers need to watch each other teach particular skills, figure

out what constitutes best practice, and work together over an extended period of time to develop consistent use of the best methods of teaching particular skills.

Frequent monitoring of the progress of every student is also as important in educating disadvantaged students as it is in taking care of cystic fibrosis patients. Making this happen requires an organization in which all adults are responsible for the well-being of all students, especially because knowledge of individual students and their families will be distributed among teachers, and this needs to be shared systematically to support the learning needs of individual children.

So how do we learn more about the factors that facilitate and those that hinder the creation of learning organizations that educate disadvantaged children well? Some important work has already been done examining, for example, responses to state accountability systems, NCLB, and collectively bargained rules concerning where teachers work and how their compensation is determined. However, little of that research has examined the impacts of public policies from the perspective of their impacts on schools in high poverty neighborhoods that are struggling to improve their performance. I think this is worth doing. Some public policy questions that are likely to affect the development of schools as learning organizations include the rules regarding the selection of teachers; how long the school has to demonstrate student learning gains before it is reconstituted or closed; whether the school is obligated to accept children in the middle of the school year; whether it has the resources to deal with emotionally disturbed children who disrupt instruction.

There is another set of questions that are even more fundamental to the efforts of educators to create learning organizations that are effective in educating disadvantaged children. They concern the curriculum, how it is taught, and what is required of students. While there is

agreement that making instruction more consistent is important, there is no agreement on just what this instruction should look like. Nor is there agreement on requirements for students – that is, the grounds on which they may be dismissed if they violate school rules. Moreover, there is substantial variation in these dimensions among schools struggling to educate disadvantaged children.

To many public policy analysts, these seem like relatively uninteresting questions that should be left to the educators. However, they have a huge impact on the work that teachers do and on the daily experiences of children in schools. School leaders make big bets on curriculum and instructional approaches and disciplinary rules, and they lack the answers to critical questions that should inform their choices.

Let me mention a few of the options, recognizing that there is variation within each option and overlap among the categories. The first is called the “no excuses” approach. It is often identified with KIPP schools and more recently, with the Promise Academy in the Harlem Children’s Zone. Common components include a longer school day and school year, a rigorous disciplinary code with rewards and sanctions, a contract with parents to support the school’s efforts, and a very large amount of time devoted to preparation for state reading and math tests. According to Paul Tough, whose book, *Whatever It Takes*, describes the Promise Academy, instruction tends to be very teacher-centered and didactic (Tough 2008).

Encouraging news about the potential for these schools to make a real difference in the lives of children living in truly disadvantaged circumstances comes from Will Dobbie’s and Roland Fryer’s paper documenting that children who won the lottery to obtain a place in the Promise Academy made dramatically larger achievement gains as measured on state tests than did children who lost out in the lottery (Dobbie and Fryer 2009).

One question about the no excuses schools is whether their strong disciplinary code and the requirement that parents sign a contract lead the most troubled students to avoid these schools, or to leave these schools after a year or two. This does not mean that the schools are not doing important work. However, it does raise the question of where the most troubled students go to school and the costs associated with educating them.

A second question is whether the improvements in scores on the state-mandated reading and math tests will translate into better long-term outcomes for the students. To my knowledge, we know little about the long-term consequences for this student population of pedagogies that emphasize test preparation. They may enable students to develop the confidence and skills to succeed in high school and college. It seems very important to learn whether this is the case.

A second, highly publicized whole school reform approach is *Success for All*, a comprehensive attempt to change reading practices in elementary schools. Elements include lesson scripts calling for fast paced, teacher-led direct instruction, and frequent administration of curriculum-aligned reading assessments, the results of which are used to regroup students by reading level. Participating schools are required to appoint a program facilitator whose duties include assuring that teachers are faithfully implementing the program (Rowan 2009). A strength of *SFA* is that the detailed scripts make the program relatively easy to implement, even with teachers who have limited skills and experience. Positive evidence about the impact of *SFA* comes from a random assignment evaluation by Geoff Borman and his colleagues showing stronger reading skills in elementary school for students in schools using *SFA* than for students in schools using other reading programs (Borman 2007).

I have asked two very knowledgeable colleagues about *SFA*. One believes that while it is effective in teaching basic reading skills, it is not effective in preparing students to make sense of

the more difficult texts they will encounter in middle school and high school. The other argues that *SFA*, when implemented well, provides a very strong foundation for later success. To my knowledge, there are no evaluations of *SFA* with long-term follow ups that would shed light on this issue. Yet hundreds of schools make decisions every year about whether to adopt *SFA*.

A third approach to instruction is to adopt curricula that specifically aim at developing students' understanding of core concepts rather than teach rules. Typically, the instructional methods associated with these curricula are student-centered and emphasize group work. An example is the *Everyday Mathematics* curriculum, which many suburban elementary schools use and a significant number of urban districts have adopted. When *Everyday Mathematics* is taught well, it is very impressive. However, doing so requires significant understanding of mathematics as well as skill in managing multiple small groups. Evidence of the potential and challenges of adopting this type of curriculum comes from an evaluation of *Everyday Mathematics* in Pittsburgh. After three years in which significant resources were devoted to developing teachers' skills in teaching this curriculum, the evaluators reported that in classrooms where it was implemented well, students had made exceptionally large gains in scores on mathematics tests. However, the report also stated that less than half of the Pittsburgh teachers were implementing *Everyday Mathematics* well after three years of coaching (Briars and Resnick 2000).

Directions for Research

So, there are quite different approaches to teaching disadvantaged children, and some evidence supporting the promise of each approach. However, there also are many unanswered questions that bear on the effectiveness of schools in educating disadvantaged children well. They include:

- Are there differences in longer term impacts – for example, in preparing students to do high school work?
- To what extent do outcomes of different approaches depend on the academic strengths of the teachers a school is able to recruit – and consequently, on the tools available to schools to recruit teaching talent?
- How long does it take with different approaches to show progress?
- Does the relative effectiveness of different approaches depend on the amount of student mobility?
- Does it depend on the extent to which schools have the right to dismiss emotionally disturbed students who frequently disrupt instruction, or on whether they have the resources needed to deal with these children?

And finally, do the results of different approaches depend on accountability rules, such as the *No Child Left Behind* rule that schools demonstrate Adequate Yearly Progress each year? Reason to ask about this is the extraordinary amount of time many high poverty schools devote to test preparation. One way to learn about this would be to allow groups of schools that had a promising comprehensive school improvement strategy for educating disadvantaged children to obtain waivers from salient provisions of *No Child Left Behind*. A condition for obtaining a waiver could be that resources be set aside for a high quality impact evaluation, one that would examine a variety of student outcomes.

There is precedent for such waivers from federal government policies. In the 1980s, before the devolution of welfare policies to states, *HEW* provided waivers from federal rules to states that wanted to innovate. A condition for a waiver was that resources be set aside for a high quality impact evaluation, usually with random assignment. Lessons from these evaluations

played a major role in informing the design of subsequent welfare reform legislation. I see similar promise for the creation of knowledge that would inform the design of future state and federal education policies.

Final Words

I started out describing the reasons why improving the education of disadvantaged children is so important for our country. I then described research evidence that, on the one hand, supports the importance of attracting teaching talent to schools serving concentrations of disadvantaged children. However, I also argued that puzzles in this research evidence suggest that educating disadvantaged children well requires more than simply attracting skilled teachers, each of whom works independently. This led me to consider a set of questions that bear on the ability of schools to become learning organizations that continually improve their performance in educating disadvantaged children. I hope to listen to presentations at future APPAM conferences that bear on these questions.

I close by suggesting three indicators of whether the nation is making progress in educating disadvantaged students: one relatively short term, one medium term, and one, longer term. The short term indicator of success would be that schools serving high concentrations of disadvantaged children have become good enough places to work that they can be quite selective in choosing among a large number of skilled applicants for teaching positions. I don't know what combination of incentives will be required to bring this about. However, I believe that it will require progress in creating schools that are learning organizations. The reason is that this is what it will take to educate disadvantaged children well, and the strongest incentive for good teachers is to see their work making a difference in the lives of children.

The medium term indicator of progress would be significant increases in the high school graduation rate of disadvantaged youth and in the rate at which they enroll in college and succeed in post-secondary education without significant periods of remediation. In a society in which education credentials play such an important role in determining labor market outcomes, this is a critical step toward a more just society.

The longer term indicator of success would be a marked increase in intergenerational economic mobility – that is, real progress toward a society in which being born into poverty did not predict nearly as well as it does now that children and their children will live their lives in poverty.

References

Bjorklund, Anders, and Markus Jantti. 2009. Intergenerational income mobility and the role of family background. In *The Oxford handbook of income inequality.*, eds. Wiemer Salvedra, Brian Nolan and Timothy M. Smeeding, 491-521. New York: Oxford University Press.

Borman, Geoffrey D. 2007. Final reading outcomes of the national randomized field trial of success for all. *American Educational Research Journal* 44, (3) (09): 701-31.

Briars, Diane J., and Lauren B. Resnick. 2000. *Standards, assessments -- what else? the essential elements of standards-based school improvement.* Pittsburgh: UCLA, National Center for Research on Evaluation, Standards, and Student Testing/University of Pittsburgh, Learning Research and Development Center, CSE Technical Report 528.

Corcoran, Sean P., William N. Evans, and Robert M. Schwab. 2004. Women, the labor market, and the declining relative quality of teachers. *Journal of Policy Analysis and Management* 23, (3) (Summer 2004): 449-70.

Decker, Paul T., Daniel P. Mayer, and Steven Glazerman. 2004. *The effects of Teach for America on students: Findings from a national evaluation.* Princeton, N.J.: Mathematica Policy Research, Inc.

Dobbie, Will, and Roland G. Jr Fryer. 2009. *Are high-quality schools enough to close the achievement gap? evidence from a bold social experiment in Harlem.* Harvard University Working Paper ed. Cambridge, MA.

Gawande, Atul. 2004. The bell curve. *The New Yorker.* December 6.

Goldin, Claudia D., and Lawrence F. Katz. 2008. *The race between education and technology.* Cambridge, Mass.: Belknap Press of Harvard University Press.

Jackson, C. Kirabo, and Elias Bruegmann. 2009. Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics* 1, (4): 85-108.

Kane, Thomas J., and Douglas O. Staiger. 2008. Estimating teacher impacts on student achievement: An experimental evaluation. National Bureau of Economic Research, Inc, NBER Working Papers.

Levy, Frank, and Richard J. Murnane. 2004. *The new division of labor: How computers are creating the next labor market.* Princeton, N.J.: Princeton University Press.

McCaffrey, Daniel F., Tim R. Sass, J. R. Lockwood, and Kata Mihaly. 2009. The inter-temporal variability of teacher effect estimates. *Education Finance and Policy* 4, (4): 572-606.

OECD. 2009. *Education at a glance 2009: OECD indicators*. Paris: OECD.

Rivkin, Steven G., Eric A. Hanushek, and John F. Kain. 2005. Teachers, schools, and academic achievement. *Econometrica* 73, (2) (March 2005): 417-58.

Rockoff, Jonah E., Brian A. Jacob, Thomas J. Kane, and Douglas O. Staiger. 2008. Can you recognize an effective teacher when you recruit one? National Bureau of Economic Research, Inc, NBER Working Papers.

Rowan, Brian. 2009. *Working in and around schools to improve the education outcomes of students in poverty: Toward a comprehensive agenda for school reform in the United States*. draft paper written for the Social Inequality and Educational Disadvantage Project ed. Ann Arbor, MI.

Tough, Paul. 2008. *Whatever it takes*. New York: Houghton Mifflin Harcourt.

Watson, Tara. 2009. Inequality and the measurement of residential segregation by income in American neighborhoods. *Review of Income and Wealth* 55, (3) (September 2009): 820-44.