“Resolved: That the federal government should equalize educational opportunity by means of grants to the states for public elementary and secondary education.”

Fifty years ago I began my work analyzing educational reform as a member of my high school debating team, focusing on this 1961-62 national topic. While I forgot many things about high school, I remember this topic because my debating partner creatively suggested that we take the position in favor of federal aid to education with an argument that federal aid would lead to federal control and that would be good. That threw our “con” opponents for a loop because the national debate briefing book was based on arguing federal control would be harmful, stifling innovation in the states and injecting federal bureaucrats into an area of state and local control. The standard “pro” defense when attacked by those against federal aid was that federal control
would not result. So here it is 50 years later and our nation is still considering how to reform public education to achieve equal educational opportunity and what the appropriate role of the federal government in education should be. So with this 50-year milestone it was time for me to look back and review what we have learned about reforming American education.

Today the goals of the American public education system have changed somewhat. Equality of opportunity has morphed into equality of results, that is reducing the racial and class achievement gaps among students, and has been joined by the goal of educating all students to a certain level (as captured by the slogan, “All children can learn.”), increasing the efficiency of providing public education, and competing successfully in the world economy. Many in our nation are concerned that while we have increased our financial commitment to education, our students have not shown progress in various assessments and are far from the top in international comparisons. As but one example, the 2009 PISA results, comparing student scores in reading, mathematics, and science across industrialized nations, placed the U.S. behind leaders South Korea, Singapore, and Shanghai and near the Organisation for Economic Co-operation and Development (OECD) mean (OECD, 2010; Robolen, 2010). These goals serve as the backdrop for any current analysis of reforming American education. The social science backdrop of reforms extends back to the mid-1960s and the Coleman Report.

Equality of Educational Opportunity

The Civil Rights Act of 1964, proposed by President Lyndon Johnson and enacted by Congress after the assassination of President Kennedy, included a provision to conduct a national study of the resources available to students across the nation and how resource inequities were
linked to inequality of educational achievement. The then princely sum of $1 million was appropriated by Congress for this task. Johns Hopkins sociologist James Coleman directed the research team. The study involved the testing of 640,000 students in a wide range of grades and surveying their 60,000 teachers in 4,000 schools (Gamoran & Long, 2006).

The so-called Coleman Report (Coleman, Campbell, et al., 1966) was released on July 4, 1966. Releasing the report on this holiday had nothing to do with patriotism but much to do with trying to downplay the results. The study found:

- Family background was the major variable explaining test score variation.
- Greater variation existed in test scores within than across schools.
- Measured school resource variation was small and not an explanation for achievement differences.
- Significant variables linked to achievement were the social background of peer groups, teacher verbal ability, and student sense of control over their fate.
- Schools were highly racially segregated.

In other words, the reasons for the vast gap between white and black student achievement could not be laid at the door of resource differentials such as years of teacher experience, availability of science labs, and funding per student, but rather were due primarily to non-school factors. This was not palatable to the left, which assumed resources would solve the problem. And the major variable that was school based, peer values, implied school desegregation would help to alleviate the racial/income gap, an anathema to the right. While social scientists criticized the methodology and reanalyzed the data (Mosteller & Moynihan, 1972), the basic conclusions
The report helped to launch the social science analysis of educational reform and shifted the goal from the equality of opportunity to the equality of results (Coleman, 1973).

50 Years of Reforms

For the past 50 years, a panoply of educational reforms have been put forward as panaceas for solving the education crisis in the nation. Figure 1 divides these alternatives into those focused on governance issues, including community control and accountability, and school-based reforms, such as compensatory programs and class size reductions. One by one these reforms have been enacted and found wanting or, at best, limited in impact. The favorite of teachers and parents is reducing class size, that is, the number of students in each classroom. Like many of the reforms, it has intuitive appeal. If there are fewer students to teach, teachers can devote more time to each student and that will increase learning. It is pure common sense! But until the Tennessee randomized experiment, where students were assigned randomly to classrooms with different class sizes, the evidence on this policy alternative was mixed at best (Whitehurst & Chingos, 2011). The Tennessee experiment gave testimony to the advocates of class size reduction for the early grades (K-3). California was then moved to implement state-based reductions in class size. However, lacking adequate classrooms and quality teachers, the reform failed (Whitehurst & Chingos, 2011). Charter schools, schools of parental choice which are not subject to some of the rules, regulations, and statutes that apply to other public schools in exchange for some type of accountability set forth in each school’s charter, were implemented with high hopes. This reform had auspicious origins, advocated on the left by American Federation of Teachers (AFT) union president Albert Shanker as a way for teachers to control schools and by the right as a market-based reform of education.
based on parental choice. Like class size reduction, this reform struck a popular chord. First enabled by a 1991 Minnesota law, today there are over 5,000 charter schools in the nation serving about 1.5 million students (http://www.edreform.com/_upload/CER_charter_numbers.pdf accessed on May 31, 2011). Despite the popularity of this reform in many circles, especially among business elites, conservatives, and African Americans, the results have been mixed. Recent studies have not found that charter schools serving students with the same backgrounds as regular public schools have consistently been any better at increasing student achievement (CREDO, 2009). Advocates are now stating that there are some excellent but also some inadequate charter schools and that we must focus on the positive. Of course, we can also say that about regular public schools. The mode of governance does not seem to be a magic bullet.

FIGURE 1

50 Years of Education Reforms: The Search for a Panacea

<table>
<thead>
<tr>
<th>Governance-based</th>
<th>School-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Larger school districts</td>
<td>Compensatory programs for disadvantaged</td>
</tr>
<tr>
<td>Community action; community control</td>
<td>School desegregation</td>
</tr>
<tr>
<td>Decentralization</td>
<td>Pedagogical models</td>
</tr>
<tr>
<td>Funding equity</td>
<td>Standards-based education</td>
</tr>
<tr>
<td>Private and parochial schools</td>
<td>Technology</td>
</tr>
<tr>
<td>Tuition vouchers</td>
<td>Preschool education</td>
</tr>
<tr>
<td>Accountability (NCLB)</td>
<td>Class size reduction</td>
</tr>
<tr>
<td>Charter schools</td>
<td>Smaller schools</td>
</tr>
</tbody>
</table>

Perhaps the Stanford historian of technology in education Larry Cuban’s book title captures not only the disappointment in technology to transform schools best, but the limits of
other reforms over the last five decades—*Oversold and Underused: Computers in the Classroom* (Kuehn, 2002). Of course, the future always looks promising—now there is the iPad II and cloud computing. In my view only pre-school education has proven to be a cost-effective strategy for improving student achievement of low-income children (see Reynolds, Temple, et al., 2011, for example), and I wonder if that is because I have not studied this literature sufficiently.

Despite 50 years of disappointment, the solutions keep coming. Recent examples touting new reforms include:

- *What’s Wrong with Our Schools and How we Can Fix Them* (Zwaagstra, Clifton, & Long, 2010);
- *Schooling at the Speed of Sound: A Blueprint for Making Schooling More Effective* (Lloyd, 2010); and

Since there is not space in this article for an analysis of every failed reform, let me address the larger question: why don’t education reforms have a greater impact? My answer begins with a return to the Coleman Report. The finding that non-school factors have a powerful role in determining student achievement has not changed over the decades. A comprehensive review of studies indicates that student characteristics such as home background and student motivation account for 80 percent of the variation in student achievement (Marzano, 2000). The remaining 20 percent is divided among school characteristics (7%) and teachers (13%). In brief the findings of the Coleman Report have held up over time. “Forty years on, the findings of the Coleman report hold up remarkably well, in some ways distressingly so...In light of these persisting patterns,
the lessons of EEO and the research that followed leave little room for optimism about the power of schools and schooling to bring about equality of opportunity in the sense of equality of results” (Gamoran & Long, 2006). Oft-quoted bank robber Willy Sutton was asked why he robbed banks and he declared, “That’s where the money is!” And the “money” or payoff in improving student achievement is not primarily in schools.

There is not much causation, at least as our society is now structured, in school characteristics for reforms to attack. As I watch the advantages my own grandchildren have – two educated parents; resources (including money and time) for preschool, music lessons, sports, scouts, and other activities; great models of hard work and moms who are voracious readers; smart and school-oriented peers in their well-off and well-educated neighborhoods; and high expectations and college visits before puberty has begun – I see how home environment plays its role. Research has carefully documented the role that home background plays. For example, Hart & Risley (1995) estimated that in a year a child in a professional family would hear 11 million words while a child in a welfare family would hear only 3 million. Home background does not guarantee success or failure but it does affect the difficulty of the path to school success. It should be noted, however, that the degree to which home background now affects student achievement is not necessarily immutable. This percentage is less in some other nations and, perhaps with appropriate policies, could be altered (Gamoran & Long, 2006).

The reader might ask, “But I am always reading about these great schools that overcome all odds and help the students to exceed expectations! Don’t we already know how to make schools work?” Perhaps the answer is it is sometimes evident that at least someone knows how to make a school work but translating the success into making many schools work is not easy. It is not clear
that there is one solution to the problems of schools in rural Iowa and inner-city Chicago. Scaling up and implementation has been studied since the 1970s and remains an issue (Murphy, 1971; Sparks, 2011). Changing the status quo is a challenge as the recent brouhaha in Delaware over the Christina School Board’s backtracking on school reforms they had adopted, specifically transferring teachers out of a failed school, were nixed once the impacts were clear and specific. Only intervention by the Secretary of Education of the nation and of Delaware led the school board to reconsider its actions (McNeil, 2011). And even when success is clear, the reasons for success may not be. Was a school’s success due to the principal, teachers, curriculum, or other factors far more subtle? Education is an arena where everyone has had personal experience and thinks they know how to improve it – usually “teach the way I was taught” – but simple, common sense answers have rarely found an empirical basis for success, which brings me to the latest reform.

**Teacher Accountability with Value-Added Measures**

As I write this article a new reform is attracting a great deal of attention. This reform is focused on teachers. Recent research has indicated that teachers may have a profound effect on student success. Talk about common sense! I am sure everyone reading this work has had a teacher to thank for motivating them to do their best, for teaching them a significant skill to move ahead, to give them a broader view of what they could do and accomplish. Indeed the Coleman Report found that the score teachers received on a short vocabulary test was related to student achievement. (But it should be noted that the cross-sectional nature of the research begged the question of whether vocabulary led to student achievement or teachers with better vocabularies
were attracted to schools with higher achieving students. And what did the vocabulary test measure—English language ability, intelligence, or an ability to master teaching skills?)

Statistician William Sanders, working in Tennessee, developed the notion of value added, determining the added value a school, and later a teacher, brought to student achievement. (See Wainer, 2004, for an excellent volume on value-added assessments.) Other researchers have found that the factors on which teacher salaries are based, years of experience and degrees and credits, are not significantly related to student achievement. Current teacher evaluation systems are obviously flawed, with one study finding that 99% of teachers were (usually perfunctorily) rated “satisfactory” and only about 1% unsatisfactory (Weisberg, Sexton, et al., 2009), again violating common sense and the experience of all those who have ever been in school. And the laying off or RIFing (reduction in force) teachers based on seniority (“last hired, first fired”) rather than effectiveness has riled many a parent. Finally, contracts that allow more senior teachers to choose to teach in schools with more advantaged students work against closing the achievement gap.

FIGURE 2

Credential-based versus Performance-based Policies

<table>
<thead>
<tr>
<th>Current policy</th>
<th>Proposed reforms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluations perfunctory and “satisfactory”</td>
<td>Evaluations based on student test scores</td>
</tr>
<tr>
<td>Tenure after 3 years</td>
<td>Tenure after 5 years, based on test scores, or tenure eliminated</td>
</tr>
<tr>
<td>Salary based on degree and experience</td>
<td>Salary based on test scores and best practice (merit pay)</td>
</tr>
<tr>
<td>Last hired; first fired</td>
<td>RIFs based on effectiveness</td>
</tr>
</tbody>
</table>
In short, reformers seek to change the major elements of the system by which teachers are rewarded. Figure 2 summarizes the current system and the reforms proposed in evaluation, the granting of tenure, salary, and RIFing. Indeed one researcher, conservative economist Eric Hanushek, estimates that eliminating the worst teachers as measured by their value-added effectiveness would net the U.S. gross national product a present value of $100 trillion (Hanushek 2010). (See how one can get carried away!)

What are value-added measures of teacher productivity? “The contribution of a teacher to student learning can be estimated by comparing the average achievement of a teacher’s students to the level of achievement that would be expected...” (Odden, 2011, 79). The notion is that we can measure a teacher’s effectiveness by how well his/her students perform on a standardized test in comparison to their expected performance and then base major personnel decisions such as salary, tenure, and even firing on this effectiveness measure.

Improving teacher effectiveness with value-added measures playing a role in education personnel decisions has become the latest panacea. Teacher effectiveness is one of the four pillars of President Obama’s Race to the Top initiative:

- Adopting standards and assessments that prepare students to succeed in college and the workplace and to compete in the global economy;
- Building data systems that measure student growth and success, and inform teachers and principals about how they can improve instruction;
- Recruiting, developing, rewarding, and retaining effective teachers and principals, especially where they are needed most; and
• Turning around our lowest-achieving schools


A major element of the effective teacher pillar is improving teacher and principal effectiveness based on performance. In part spurred by RTTT and in part by the election of more conservative Republican governors, using value-added measures of teacher effectiveness has become the cornerstone of proposed educational reforms in many states (Cavanagh, 2011). Figure 3 lists a subset of these. Many of the reforms propose merit pay based on value-added measures. In the most extreme of these changes, Ohio has replaced salary schedules and step increases with mandatory merit pay for teachers with half of their rating based on one year’s student growth on standardized tests.

**FIGURE 3**

**Recent State Reforms of Teacher Policies related to Value-Added Measures**

<table>
<thead>
<tr>
<th>“Fifty percent of a teacher’s evaluation should be based on direct measures of student achievement as demonstrated by assessments and other evaluations of student work.” (New Jersey Task Force, 2011)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teachers in Tennessee will now have to go through a five year probationary period (up from three years) and have a documented impact on student performance in order to obtain tenure.</td>
</tr>
<tr>
<td>The Washington Senate approved H.B. 1443 that would put teachers who score lowest on performance evaluations the first in line for layoffs.</td>
</tr>
<tr>
<td>Virginia has competitive grant program for hard-to-staff schools for teachers, who earn exemplary ratings where at least 40% of rating determined by student growth measure. (up to $5,000 per teacher)</td>
</tr>
<tr>
<td>Ohio just replaced salary schedules and step increases with mandatory merit pay for teachers with 50% of rating based on one-years’ student growth on standardized test.</td>
</tr>
</tbody>
</table>
Unfortunately this rush to reform has underplayed or ignored many technical and policy issues that must be addressed with value-added measures. Harris’ new book (2011) on value-added measurement of teacher effectiveness does an excellent job of discussing the technical issues including:

- Inadequate data systems (not all states or districts have the data needed to create a value-added measure);
- No agreed-upon method (the determination of a teacher’s value-added contribution is in its infancy);
- Measurement error (reliability);
- Instability of categorization (a teacher’s level of effectiveness may well differ by the method used and from year to year);
- Untested grades and subjects (how do you include teachers in subject areas without standardized tests such as kindergarten, art, and physical education?);
- Validity of student assessments (e.g. to students the assessments are low-stakes tests, that is, they have modest incentives to do well); and
- Inadequate comparisons (e.g., it is more problematic to make comparisons in smaller districts).

The technical issues may well be addressed over time but the policy issues are too easily glossed over. Standardized test scores are an important measure of student learning but how important are they in comparison to motivation, socialization, citizenship, leadership, and values? Nations which have focused on rote learning for exams, such as high-scoring and leader in
economic development South Korea, are trying to copy American education to develop more creative and innovative graduates (Rhodes, 2010). Another policy issue is to what extent should the expected level of student performance take into account race and social background? Some could reasonably argue that these should play no role because we do not want to lower expectations for any particular group. Others reasonably argue that not to incorporate such factors creates unreasonable expectations, pushing teachers away from taking on challenging students.

Merit pay seems to appear on the educational reform radar scope about every 20 years and studies have not shown that its increases student achievement. (See Springer, 2009, for an excellent collection of papers on this topic.) One reason is that teachers tend to be motivated by intrinsic not extrinsic factors. To quote a principal in a study we recently completed in Delaware, “Elementary school teachers teach because they love kids. Secondary school teachers teach because they love their subject. And college teachers teach because they love themselves” (Farley-Ripple, Mead, et al., 2011). Ignoring the last obviously gratuitous comment, one could argue that a merit pay system based on teacher performance might bring a different type of teacher into the field. But the evidence would suggest the opposite—the most successful program attracting America’s most talented young people into teaching, Teach for America, is based on intrinsic, not extrinsic, motivation.

Despite these issues, teacher value-added proposals linked to hiring, tenure, evaluation of teacher preparation institutions, salary, tenure, firing, and RIFing are being generated almost daily. Why are such proposals so attractive? I believe there are several reasons. First, these proposals have public support. In a recent Education Next poll half the pubic favored merit pay for teachers and only one-quarter opposed it (Howell, Peterson, & West, 2010). Most people work in situations
where their productivity is related to salary and thus personal experience and common sense would support this idea. Second, value-added proposals are inexpensive and often revenue neutral since there is a substitution of current determinants of salary for value-added effectiveness measures (e.g., Ohio’s new legislation). In an era of budget constraint, this is attractive and feasible. Third, these proposals do not upset the status quo of sorting the better off from the poor, i.e. they are non-redistributive, thus avoiding major opposition from powerful groups. And last but not least, these proposals offer a scapegoat. Teachers are to blame for the nation’s public education failures. Governor (and University of Delaware graduate) Chris Christie of New Jersey is perhaps the most vocal adherent of this position.

So will value-added measures of teacher effectiveness forming the basis of personnel decisions decrease the achievement gap, ensure that all children can learn, increase educational efficiency, and increase America’s standing in international assessments? Or will this wave of reform when put to the test of actual implementation and program evaluation fail to alter the pattern of unsuccessful reforms of American education for the last 50 years?

There is some room for guarded optimism here. First, the positive effect of good teachers on student achievement not only makes common sense, there is empirical evidence, from the Coleman Report to recent studies cited above, that some teachers produce better results than others. Increasing our consciousness about which teachers achieve more with students could help us to better evaluate and improve teacher preparation programs, professional development efforts, teacher recruitment and selection processes, and other components of the human capital system in education. However, I would add that there are many less problematic improvements we could
make in the human capital system which would help, and we need to be somewhat skeptical about the extent to which changing the current incentive system will increase achievement.

Here are some human capital improvements of which I am aware in Delaware. For almost the last decade work I have conducted for the Delaware Department of Education through the Institute for Public Administration has documented the hiring of a majority of new teachers in Delaware in August or later. For example, 64 percent of teachers receiving regular contracts were hired in August or later for the fall of 2010 school year, and an additional 400 plus teachers were hired on temporary contracts, mostly because they, too, were hired late, if not even after the school year began (Raffel & Cox, forthcoming). This late hiring has several downsides including a lack of time for teachers to prepare for the start of their first teaching assignment, the negative signal it sends teachers about their profession and their position, and the question of whether the state has an opportunity to hire the most competitive candidates for teaching positions. The major (but not only) reason for this late hiring has been uncertainty over the September 30 enrollment count and the state funding allocation based on the count. The Delaware General Assembly recently passed Senate Bill 16, which will establish a one-year trial of estimating enrollment in April and guaranteeing school districts 98 percent of the funding based on those enrollments.

A 2010 follow-up to a 2004 survey of Delaware’s first-year teachers I conducted indicated those who had a positive mentoring experience were more likely to remain as teachers in the state six years later. The Delaware New Teacher Mentoring and Induction program, which requires new teachers’ participation for three years, has recently been the subject of budget cuts and loss of key personnel. The 2010 survey of teachers in their first three years indicated a precipitous drop in satisfaction with this program, suggesting a less effective program and a subsequent increase in
teacher attrition (Raffel & Welch, 2010). Given the potential of such induction programs, the disinvestment in the program was not a wise move.

The expectations for changes in personnel policies based on teacher value-added measurement, as well as many other reforms including most significantly charter schools, are based on an economic model of motivation. At the heart of this model are the values of competition, choice, and extrinsic motivation. While this model is useful in understanding and changing some consumer behaviors, the values which motivate almost all teachers are more likely to be collaboration, child-centeredness, and more generally intrinsic motivation. Linda Darling Hammond’s work is the best exemplification of viewing educational reform through this professional as opposed to economic lens (for example, see Darling-Hammond, 2010).

Unfortunately the current effort to reform schools based on teacher productivity and the economic model and often focused on disincentives may backfire. Nations which have succeeded in changing their educational success rate and stand at the top of international assessments, Singapore, Korea, and Finland, hold teaching in the highest regard and invest heavily in teacher recruitment and selection, development, and support (Alliance for Excellent Education, 2011). That is the opposite of the current reform wave as governors and others are attacking teachers and their unions. We may wind up rewarding the best teachers of today but fail to recruit those who would be the best teachers tomorrow. Indeed, Marc Tucker, National Center on Education and the Economy, recently concluded that the U.S. needs to go in the opposite direction, not belittling teachers but boosting them:

No nation can move the vast majority of students to the levels of intellectual capacity and creativity now demanded on a national scale unless that nation is recruiting most of its
teachers from the group of young people who are now typically going into the non-feminized professions. Recruiting from that pool requires a nation not just to offer competitive compensation but also to offer the same status in the society that the non-feminized occupations offer, the same quality of professional training and the same conditions of work in the workplace (Tucker, 2011, 3).

Final Reflections

Looking back over 50 years of attempts to reform American education, I am struck by how the results of the Coleman Report have remained relevant and significant. One implication is that meeting America’s education goals must be addressed in large part outside of American schools. Given that most of the explanation for variation in achievement is based on factors external to schools, we must include addressing external issues as part of any attempt to move forward (Henig & Reville, 2011). Indeed, in 2008, Helen Ladd, a Duke University professor and education researcher, along with educators such as U.S. Secretary of Education Arne Duncan, and former Boston school superintendent Dr. Thomas Payzant, and many well-known social scientists and think-tankers started an organization and movement to do just that—the Broader, Bolder Approach to Education campaign—to focus on the well-being of disadvantaged children (kidshttp://www.boldapproach.org/who-we-are accessed on May 31, 2011). The nation must address the child poverty rate, single-parent families, and increasingly diverse and disadvantaged student body and devote more resources to our nation’s children (Kirp, 2011).

Second, the role of social science, both in terms of concepts and research, has become increasingly important. There are many common sense solutions to the problems of American
education that when implemented do not work. (Recall the quote: “For every complex problem, there is a solution that is simple, neat, and wrong.” -- H. L. Mencken.) Understanding the theories, concepts, and models which underlie these reforms help us to grasp what may lead to failure or success.

Finally, it is important to recognize what we have accomplished in these 50 years. We have far better data on not only what our nation’s students are achieving through the National Assessment of Educational Progress (NAEP), we also have international comparisons through PISA and TIMMS. We have a better understanding of why social background matters. And perhaps most of all, we are no longer debating the significance of education on the state or federal agenda but how investments should be made.
References


