

## **Alice in Wonderland Fantasies: Many National Educational Reform Models: Not Based on Research**

**Author(s):** *Arthur Shapiro*

**Affiliation:** University of South Florida

2011

Where once I had been hopeful, even enthusiastic, about the potential benefits of testing, accountability, choice, and markets, I now found myself experiencing profound doubts about these same ideas....my views changed as I saw how these ideas were working out in reality.

Diane Ravitch, *The Death and Life of the Great American School System*

The following comprise several major centralizing politically- and ideologically-driven national educational reform models in the United States:

- No Child Left Behind (NCLB) and that Holy Grail, Accountability
- Evaluating teacher quality by student test scores
- Tying incentives to test scores
- Improving student achievement by standardized testing
- Merit pay
- Charter schools
- Grading schools: A (high) to F (failing)
- Failing kids
- National "common core" curriculum standards

The problem? Simple. None of these panacean reform efforts are supported by research. None. Are all fantasies?

### **No Child Left Behind**

At an epochal conservative American Enterprise Institute conference on NCLB on Oct. 30, 2006, Ravitch (2010) reported presenters asked whether "... major remedies prescribed by NCLB especially choice and after school tutoring were effective....The various presentations...demonstrated that none of the prescribed remedies was making a difference....Choice was not working they all agreed....choice had little or no effect on school achievement" (p. 99), and was used only by two percent of eligible students (Bracey, 2007, February). "Free after-school tutoring (called Supplementary Educational Service, or SES) fared only a bit better than choice" (Ravitch, p. 100).

The other two NCLB tools, corrective action and restructuring, rarely involved state takeover of schools. "...Henig noted that for political scientists, policies imply theories. However, he could not identify any theory implied by SES" (Bracey, 2007, February, p. 475).

Thus, noted Ravitch, formerly a strong supporter of choice, accountability, testing and markets, "...that was the day I realized NCLB was a failure" (p. 99). Compounding this miasma are two more NCLB components: Adequate Yearly Progress (AYP) and the goal that *all* students will be at 100 percent proficiency by 2014, manifestly unrealistic and, an oxymoron. AYP is leading to huge numbers of schools labeled 'in need of improvement'. By 2014, one was noted that 70 percent of Minnesota's school will be failing. At the conference's end, Ravitch asked to much laughter "What reason do we have to believe that Congress knows how to fix the nation's schools?" (Bracey, 2007, February).

## **Accountability**

The journey for accountability appears similar to the quest for the Holy Grail depicted in Monte Pythons's King Arthur. It appears illusory since the American version of accountability is based on using test scores to determine effectiveness. Test scores obviously are unstable, prone to errors in development and particularly grading. (But, the testing industry certainly is profitable). Ravitch (2010, p. 152) noted "... standardized tests are not precise instruments", but elected officials and the public view them as rock solid, valid. Using test scores to determine accountability leads us into a squishy and alternate state of reality. We are making major decisions about kids' and teachers' and administrators' lives based often on quicksandy fantasies (as in Alice's Wonderland).

## **Evaluating Teacher Quality by Student Test Scores**

In a letter to the New York State Board of Regents (Strauss, 2011, May), 10 assessment experts comprising the Board on Testing Assessment of the National Research Council of the National Academy of Sciences stated,

Reviews of research on value-added methodologies for estimating teacher "effects" based on student test scores have concluded that these measures are too unstable and too vulnerable to many sources of error to be used as a major part of teacher evaluation.

Some sources of error include, "Teacher ratings are affected by differences in students who are assigned to them." Further, "It is impossible to fully separate out the influences of students' other teachers, as well as school and home conditions, or their apparent learning."

## **Tying Incentives to Test Scores**

Ravitch noted (2011, June), the above panel "...concluded that tying incentives to test scores has not improved education in the United States. The gains have been small to none, and the negative consequences include inflated test scores and gaming the system." Sparks (2011, May) reported similar conclusions, "Nearly a decade of America's test-based accountability systems, from 'adequate yearly progress' to high school exit exams, has shown little or no positive effect overall on learning and insufficient safeguards against gaming the system..."

In my classes this week, a sharp kindergarten teacher who loves her English Language deficient kids, told her principal she would not teach these classes again because she wouldn't base her salary on the impossibility of achieving sizeable gains with her charges. As soon as a child improved they pulled him/her into a higher class. Teachers are bailing out of working with kids who need the most help. The same problem arises with those teaching the gifted because few will "show large improvement."

## **Improving Student Achievement by Standardized Testing**

If only tests and kids and teachers and testing companies were perfect. But, they're not. So, we have imperfect instruments supposed to evaluate teachers' performance. Sparks' article's headline summarizes this hope stating forlornly, "Panel Finds Few Learning Gains from Testing Movement." Hout and Elliot (2011) noted, "Nearly a decade of America's test-based accountability systems, from 'adequate yearly progress' to high school exit exams has shown little to no positive effect overall on learning and insufficient safeguards against gaming the system, a blue ribbon committee of the National Academies of Science concludes in a new report."

## **Merit Pay**

When I was a suburban school superintendent (although I taught and administered in inner city, well-to-do suburbs, and rural settings), several tried merit. Where such a practice should work, it bombed. Now, it returns -- it will bomb again. Why?

Easy. Virtually everyone feels he/she is well above average. Indeed, in a large study of 1,088 professional and managerial employees, when asked informally and confidentially, 47 percent rated themselves in the *top 5 percent* of "others doing similar work" and another 36 percent rated themselves in

the *upper 10 percent* (Meyer, 1975). So, 83 percent perceive ourselves in these top two categories. Thus, when supervisors rate teachers (and principals) as average and below average, those rated lower than their self-rating (the majority of us), feel their self-esteem threatened, become outraged. Two consequences become inevitable: First, the teacher depreciates the supervisor's judgment. The second is to depreciate the value of teaching or the school or district.

But, the huge assumption and question is does merit pay improve student scores? Alas, no, it does not. Vanderbilt University studied merit in Nashville, TN schools finding no improvement where teachers received \$5,000, \$10,000 and \$15,000 over teachers receiving no bonus (Springer 2010). This practice generates major negatives, since in competing for merit, teamwork and collaboration suffer, as does trust, an essential ingredient enabling people to work together. A seminal study (Bryk and Schneider 2002) found that schools with low levels of trust had a one in seven chance (14 percent) of improving reading and math scores, while those with high levels had a one in two probability. No trust, no gain.

### **More Systemic Difficulties in Evaluating Teachers Accurately and Reliably**

In "Hurdles Arise in Rising Effort to Rate Teachers" (Otterman, 2010, Dec. 26), the uncertainties regarding validity and reliability surfaced.

The most extensive independent study of New York's teacher rankings found similar variability.

In math, about a quarter of the lowest-ranking teachers in 2007 ended up among the highest-ranking teachers in 2008. In English, most low performers in 2007 ended up among the highest-ranking teachers the next year... (Corcoran, 2010, Sept.)

If evaluation systems cannot evaluate teachers validly and reliably, they are worthless.

### **Charters**

And another highly touted panacea bites the dust. In a huge Stanford Center for Research in Educational Outcomes (CREDO) study 37 percent of charters performed "significantly below" local public school students, 17 percent performed better than their matched local public schools, while 46 percent were the same (Raymond, 2009). In 2007, NAEP test scores indicated that charter students scored lower than public school kids in reading and math in fourth grade and math in eighth grade, but were even in eighth grade reading (Robelen, 2008, May). The CREDO study found "wide variance in the quality of...charter schools with, in the aggregate, students...not faring as well as students in traditional public schools." Berliner (1995) noted many strategies to improve public schools comprise efforts to privatize them, to siphon funds, to depreciate them.

### **Grading Schools: A (High) to F (Failing)**

Critics term this "naming and shaming". It describes a school socio-economically. A valid evaluation by a regional evaluating body takes fifteen or more professionals three days.

### **Flunking Kids**

Another disastrous practice leading to horrendous consequences. Sixty-five percent of kids who fail once drop out. Of those who fail twice, 90 percent do not finish.

### **National "Common Core" Curriculum Standards (CCSS)**

This represents an idea looking for research support. Unfortunately, Tienken (2011, Winter) noted, "The vendors of CCSS have a problem. They have no data that demonstrates the validity of the standards as a vehicle to build 21<sup>st</sup> century skills....Even worse, standardization assumes that children are not active constructors of meaning, that they bring prior knowledge and experience to the learning situation". In short, constructivist principles regarding how people learn are violated.

### **At Last, Some Conclusions, Even Comments**

Comparing our school reform strategies with those of the world's top performing nations is highly instructive, as the National Center on Education and the Economy has done. None have fallen for our ideologically-driven testing-students-every-year mania, which the government supports. Any physician flailing about for such treatments would lose his/her license. Instead, right-wing conservatives are determined to control public education with faddish tools of market-based accountability, choice, and above all, testing, despite examples of highest performing nations heading in entirely opposite directions involving trusting and supporting teachers. An underlying purpose is to destroy teacher unions who have strongly supported Democratic candidates. Conservatives, while espousing limited government, are hypocritically centralizing and privatizing public education.

### **References**

- Berliner, D. C. & Biddle, B. J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. Jackson, TN: Perseus.
- Bracey, G. W. (2007, February). Things fall apart: NCLB self-destructs. *Phi Delta Kappan*, 88(6), 475-476.
- Bryk, A. S. & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. NY: Russell Sage Foundation.
- Corcoran, S. P. (2010, Sept.) Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice. Providence, RI: *Annenburg Institute for School Reform*. Brown University.
- Hout, M & Elliott, S. W. (Eds.). (2011). *Incentives and Test-based Accountability in Education*. Committee on Incentives and Test-Based Accountability in Public Education. Washington, D. C.: National Research Council.
- Meyers, H. H. (1975, winter). The pay for performance dilemma. *Organizational Dynamics* 3. American Management Associations. 71-78.
- Otterman, S. (2010, Dec. 26). Hurdles emerge in rising effort to rate teachers. *NY Times*, p. A1.
- Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. NY: Perseus.
- Ravitch, D. (2011, June). A standardized path to school ruin. *St. Petersburg Times*, pp. P1, P5.
- Raymond, M. E. (2009). Multiple choice: Charter school performance in 16 states. Palo Alto, CA: *Center for Research on Educational Outcomes*, Stanford University.
- Robelen, E. W. (2008, May). NAEP gap continuing for charters: Sector's scores lag in three out of four main categories. *Education Week* 27(38), 1, 14.
- Sparks, S. D. (2011, May 11). Panel finds few learning gains from testing movement. *Education Week*. <http://www.edweek.org/ew/articles/2011/05/26/33academy.h30.html?tkn=000Fdh5WFx3FP%2BD8YjRj8tB%2FX2lbo4Wt%2BjM4&print=1>
- Springer, M. (2010). Project on incentives in teaching. Nashville, TN: *National Center on Performance Incentives*, Vanderbilt University.
- Strauss, V. (2011, May). The letter from assessment experts the N. Y. Regents ignored. *Washington Post Answer Sheet*.
- Tienken, C. H. (2011, Winter). Common core state standards: An example of data-less decision-making. *AASA Journal of Scholarship and Practice* 7(4).