

FairTest

National Center for Fair & Open Testing

Paying Teachers for Student Test Scores Damages Schools and Undermines Learning

Independent researchers have found that evaluating and paying teachers for test scores is either damaging or irrelevant to improved learning. Unfortunately, even evidence of harm does not seem to affect the growing popularity of such schemes. Policymakers, including U.S. Education Secretary Arne Duncan, should stop promoting this failed approach.*

Paying for higher test scores results in score inflation, not genuine learning. Researchers have extensively documented test score inflation under the widely criticized No Child Left Behind (NCLB) and similar state programs (Koretz, 2009; Madaus, Russell & Higgins, 2009; Nichols & Berliner, 2007). Paying Wall Street executives for short-term monetary results contributed to the financial bubble and its subsequent collapse. Similarly, payment for test scores will intensify the existing trend toward inflating results (the bubble, caused by extensive teaching to the test) without improving real learning (the collapse, visible when students enter college or the workforce unprepared).

Payment for "performance" will exacerbate damage to the curriculum caused by NCLB. Schools add time and effort to the tested subjects of math and reading, then subtract it from other important areas including social studies, science, art, music, and gym. If all teachers are paid for the school's math and reading scores, this will further narrow the curriculum. In Texas, for example, science and history teachers have been forced to teach math and reading. (See McMurrer, 2007; Au, 2007; Morton & Dalton, 2007.)

It is unfair and ineffective to pay teachers for test results often marred by scoring and other errors. Testing companies are reaping huge profits but scrambling to manage the testing explosion triggered by NCLB. This has resulted in increased scoring errors and test construction problems (Rhoades & Madaus, 2003). Human scorers are often low-paid and poorly trained temporary workers who spend a few minutes making decisions with far-reaching effects for students and teachers (Farley, 2009). Students and teachers should not have their fates determined by flawed and unreliable data.

Payment for gains in student scores does not solve the problem of test-induced educational damage. There are too many flaws in "growth" or "value added" mechanisms to trust the results. Researchers at RAND concluded that "the research base is currently insufficient to support the use of [value-added methods] for high-stakes decisions about individual teachers or schools" (McCaffrey, et al., 2005; see also, Bracey, 2007). A National Research Council (2009) evaluation came to the same conclusion.

Payment for test scores may not even to raise student scores. A study in Portugal found that a focus on evaluating individual teacher performance caused a significant decline in student scores on the national exams (Martins, 2009). An extensive payment for scores program in low-income Texas districts failed to produce gains in student scores on the state test (Springer, Podgursky & Lewis, 2009). This is despite the extensive evidence of score inflation from teaching the test.

Most teachers' primary motivation is not high pay. If it were, they would have chosen another profession. Teachers know test scores are a poor barometer of their abilities, so pay for performance damages rather than enhances their sense of professionalism and morale (Whitford & Jones, 2000; Nichols & Berliner, 2007). It can decrease motivation (Ryan and LaGuardia, 1999). Payment for "performance" also has been shown to increase cheating (Pfeffer, 2007).

Paying individual teachers for student scores encourages unhealthy competition.

Cooperation among educators is critical to school improvement and student growth, as it is in many fields (Pfeffer, 2007), but paying individual teachers for score gains can reduce cooperation as teachers compete for limited bonuses (MacInnes, 2009). Even paying bonuses to schools can cause divisions among staff and with parents (*FairTest Examiner*, 1997).

Pay for "performance" causes goal distortion in other occupations in both public and private sectors. For example, when Medicare tried to evaluate providers based on the mortality rates of their open heart surgeries, some refused to operate on the sickest patients. The U.S. General Accounting Office reviewed the system and concluded that it encouraged organizations to focus resources on that which is measured: "Areas that are not highlighted in report cards will be ignored." In part because of its poor track record, payment for results is now rare in most professions. It remains common in real estate and finance (two economic bubble areas), but even there does not produce improved results (Adams, Heywood & Rothstein, 2009).

Overall, research on pay for performance finds that it rests on dubious assumptions and lacks evidence it succeeds, while there is good evidence it often fails. In testimony to the U.S. House of Representatives, Pfeffer (2007) reported that pay for performance systems often "effectively motivate the *wrong* behavior," increased pay differentiation "lowers performance," and the schemes eat up management time while they "make everybody unhappy."

Secretary Duncan has defended payment for scores requirements in the "Race to the Top" federal fund, claiming it will only be one "significant" part of the overall evaluation of teachers. However, he has proposed no limits on the weight to be given to test scores, nor has he adequately considered the evidence summarized in this fact sheet.

While payment for test-scores must be stopped, improving the quality of teaching is central to genuine school reform. This requires ongoing professional learning and development of a well-designed evaluation system whose primary purpose is to improve teaching, but which would address chronically inadequate teachers. Student learning would be included in evaluations, but test scores would be only one small part of multiple indicators of student learning, which in turn should be one component of teacher (or principal) evaluation.

***Note:** We focus on payment, but the same issues pertain to teacher evaluation, granting tenure, or dismissal. These points generally also pertain to principals and other educators.

References

- Adams, S.J., Heywood, J.S., and Rothstein, R. 2009. *Teachers, Performance Pay and Accountability: What Education Should Learn from Other Sectors*. Economic Policy Institute.
- Au, W. 2007. High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis. *Educational Researcher*, Vol. 36, No. 5, 258-267.
- Bracey, J. 2007. Evaluating Value Added. *FairTest Examiner*, July. <http://www.fairtest.org/whats-value-growth-measures>
- FairTest Examiner*. 1997. Kentucky's Assessment System. (October.) <http://www.fairtest.org/kentuckys-assessment-program>
- Koretz, D. April 29, 2009. What's Missing in Obama's Education Plan? *Education Week*. http://www.edweek.org/ew/articles/2009/04/29/30koretz_ep.h28.html?tkn=OTLFEqvaUfgkzI4vRyp6Q0c2kzhDTpngNM9B&print=1
- MacInnes, G. 2009. Eight Reasons Not to Tie Teacher Pay to Standardized Test Results. Century Foundation Issue Brief. <http://www.tcf.org/publications/education/gordon%20brief.pdf>
- Madaus, G., Russell, M., and Higgins, J. 2009. *The Paradoxes of High Stakes Testing*. Charlotte, NC: Information Age Press.
- Martins, P. March 2009. Individual Teacher Incentives, Student Achievement and Grade Inflation. Queen Mary, University of London, CEG-IST and IZA, Discussion Paper No. 4051.
- McCaffrey, D., Koretz, D., Lockwood, J.R., and Hamilton, L. 2005. Evaluating Value-Added Models for Teacher Accountability. Santa Monica: RAND Corporation.
- McMurrer, J. 2007. Choices, Changes, and Challenges: Curriculum and Instruction in the NCLB Era. Center on Education Policy. <http://www.cep-dc.org/>
- Morton, B. & Dalton, B. 2007. Changes in Instructional Hours in Four Subjects by Public School Teachers of Grades 1 Through 4 (Issue Brief). National Center for Education Statistics. <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007305>
- National Research Council, Board on Testing and Assessment. 2009. Letter Report to the U.S. Department of Education on the Race to the Top Fund. National Academy of Sciences, available at http://www.nap.edu/catalog.php?record_id=12780.
- Nichols, S.L, and Berliner, D.C. 2007. *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*. Cambridge: Harvard Education Press.
- Pfeffer, J. 2007. Testimony to the U.S. House of Representatives. <http://federalworkforce.oversight.house.gov/documents/20070313111150-45256.pdf> Rhoades, K. and Madaus, G., 2003. *Errors in Standardized Tests: A Systemic Problem*. Boston College. <http://www.bc.edu/nbctpp>
- Ryan, R. M., and La Guardia, J. G. 1999. Achievement Motivation within a Pressured Society: Intrinsic and Extrinsic Motivations to Learn and the Politics of School Reform. In T. Urdan (Ed.) *Advances in Motivation and Achievement*. Vol 11. Greenwich, CT: JAI Press.
- Springer, M., Podgursky, M., Lewis, J., et al. 2009. Texas Educator Excellence Grant (TEEG) Program: Year Two Evaluation Report. http://www.performanceincentives.org/ncpi_publications/policybriefs.asp
- Whitford, B.L., and Jones, K. 2000. *Accountability, Assessment, and Teacher Commitment*. Albany: SUNY Press.