NATIONAL CENTER ON
Performance Incentives

Research Brief

Holding Accountability to Account

July 2008

In "Holding Accountability to Account: How Scholarship and Experience in Other Fields Inform Exploration of Performance Incentives in Education"— a paper presented at the National Center on Performance Incentives research to policy conference in February — Richard Rothstein, a research associate at the Economic Policy Institute, argues educational policy makers are not sufficiently aware of the costs and benefits of performance incentive systems. As a result, they are ill equipped to evaluate the potential value of such systems.

Rothstein contends that many of today's public education challenges have been encountered in other fields. Performance incentives in particular have been analyzed by economists, business management theorists, sociologists, and historians. As a result, valuable lessons and insights can be gleaned from these fields. In particular, Rothstein finds that while the use of performance incentives is on the rise in the private sector, there is a corresponding decrease in the use of solely quantitative measures of performance as a determinant of incentive pay. Goal distortion and gaming are inevitable results when rewards are based too heavily on quantitative measures. Rothstein concludes that this reality has eluded education policy makers. Indeed, as a result of relying on quantitative measures for determining performance incentives, educators engage in what Rothstein characterizes as three common distortions: "mismeasurement" of outputs, "mismeasurement" of inputs, and reliance upon untrustworthy statistics.

"Mismeasurement" of Outputs

Rothstein contends that conventional definitions and measurements of educational outputs (typically student achievement test results) are so oversimplified that they cannot support valid accountability or performance incentive systems. Further, the incentive effect of such measures often results in goal distortion or harmful redirection of teachers' professional practice.

Goal Distortion

Under current accountability pressures to use test scores as a measure of effectiveness, many schools are refocusing resources on reading and math, the two more easily tested and quantifiable curricular areas. This has spurred a narrowing of the curriculum wherein attention is drawn away from non-tested areas (such as art, music, science, social studies, and physical education) as well as more qualitative outcomes such as discipline, cooperation, and character.

Harmful Redirection of Teachers' Professional Practice

Accountability standards can also result in a tendency for teachers to shift their attention to specific groups of students. For example, NCLB requires that each state establish a minimum proficiency level on its standardized tests of math and reading. This approach has created incentives for teachers to narrow their instructional effort to focus on students just below the proficiency point, as they have the best chance of crossing the proficiency benchmark, thereby raising results. This approach often comes at the expense of those students who are already above the target, as well as those who are far below it.

Another redirection of practice occurs when states have the option to set their own standards for academic proficiency. In such cases, thresholds for proficiency are sometimes lowered to improve perceived performance.

Rothstein argues that these distortions result from a misunderstanding of the incentive effects of commonly used measures of education output. Each of these distortions has been documented in other fields (health care, job training, welfare policy, crime control, and the private sector) and awareness of these dangers has influenced policy in other fields to a greater extent than in education.

"Mismeasurement" of Inputs

Rothstein notes that the term "inputs" in education typically refers to school resources such as teachers, class sizes, and curricula. He suggests, however, that student demographic characteristics must also be included in the definition of inputs since students' risk of failure varies by background characteristics. Though many agree with this rationale, these characteristics have proven more difficult to measure than anticipated.

Imprecise Subgroup Definitions

Under current accountability systems in public education, student performance is reported separately by subgroup based on ethnic origin (White, Black, or Hispanic) and economic circumstance (eligibility for free or reduced-price lunch). These criteria are gross and imprecise definitions of subgroups as they do not take into account the range of student background characteristics known to impact academic achievement outcomes. Rothstein argues that more sophisticated controls are required in order to develop and monitor reasonable expectations for group or individual student performance.

Risk Adjustment

The likelihood that students will meet proficiency standards varies according to their background characteristics. Accordingly, some question whether it is appropriate to hold teachers and schools serving more disadvantaged students to the same absolute standards as those serving fewer traditionally low-performing students. Should the former group instead be judged by the achievement growth of their students? And if broad subgroup definitions do not capture the considerable variation in student background characteristics, do accountability systems create incentives for cream-skimming?

Cream-skimming

Cream-skimming refers to the practice of selecting those students from subgroups who are easiest to serve and most likely to meet established performance targets. Since only gross controls for background characteristics (i.e., race and reduced-price lunch eligibility) are currently available, some schools and teachers meet expectations by subtly selecting and measuring the progress of only the least at-risk students in the subgroups, thereby distorting the overall accountability results.

Even if controls for input or output "mismeasurement" were available, Rothstein suggests that statistical analysis of test results as a means of measuring academic performance can also undermine the credibility of any high-stakes accountability systems. These "mismeasurements" have also been documented in other fields, and influenced those policy arenas to a greater extent than in education.

Effects of Untrustworthy Statistics

Despite the opinion that quantitative results provide more scientific, statistical calculations of performance measures, they are still subject to limitations such as data reliability, sampling corruption, and other forms of gaming.

Limited Data Reliability

Poor data reliability has been an impediment to the development of performance incentive systems in education. Sample sizes are small (classes for teacher accountability; cohorts for school accountability) and results are often tabulated based on a single test score. These conditions can contribute to inaccurate statistical results, as small samples are not always representative of the student population, and random external events can influence test-taking conditions in a single event. Attempts to hold schools or teachers accountable for value-added, or score gains over time, only exacerbate reliability problems because they compound errors in the beginning and ending test scores.

Sampling Corruption

Sampling corruption occurs when effort is intensified just before the cut-off point for measuring performance, leading to a result that does not truly reflect ongoing performance. Teachers and schools can overemphasize skills needed to answer test questions. "Teaching to the test" corrupts the representation of results as teachers focus their instruction only on items that are expected to be on the test rather than covering a more comprehensive curriculum, or when students are drilled immediately prior to a test in a way that is inconsistent with strategies to encourage learning retention.

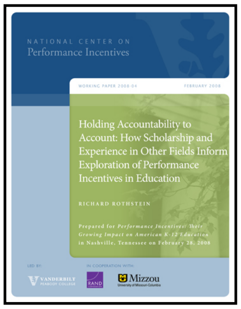
Other Forms of Gaming

Rothstein lists other forms of gaming that manipulate accountability data used to measure performance, including retaining greater numbers of low-performing students in grades prior to those being tested; excluding likely poor performers by encouraging their absence on test day or even by suspending them for real or alleged infractions; and opportunistically re-assigning students to or from subgroups (special education, English language learning, or regular education) where they can aid or do the least harm to the achievement results of sub-group performance targets.

Conclusion

According to Rothstein, the challenges associated with performance incentive systems should not come as a surprise to many education policy makers. In the private sector, performance incentive systems are used as a motivational tool, yet for the reasons cited above, professional performance awards are almost never based exclusively on quantitative measures of performance. These challenges are well documented in the research literature from other fields outside of education.

Rothstein concludes that most proponents of performance incentives and accountability systems are unaware of the extensive literature in economics and management theory that documents the distorted practices that can ensue from a heavy reliance on quantitative measures of performance, especially when used as determinants of performance incentive pay. Rothstein further contends that, without understanding of this literature, proponents of performance incentives in education are unable to engage in careful deliberation about whether the benefits are worth the price.



NATIONAL CENTER ON
Performance Incentives

WORKING PAPER 2008-04 FEBRUARY 2008

Holding Accountability to Account: How Scholarship and Experience in Other Fields Inform Exploration of Performance Incentives in Education

RICHARD ROTHSTEIN

Prepared for *Performance Incentives: Their Growing Impact on American K-12 Education* in Nashville, Tennessee on February 28, 2008

LED BY: VANDERBILT PEABODY COLLEGE

IN COOPERATION WITH: RAND Mizzou University of Missouri-Columbia

This research brief describes work published by the National Center on Performance Incentives in "Holding Accountability to Account: How Scholarship and Experience Inform Exploration of Performance Incentives in Education," by Richard Rothstein, Working Paper 2008-04. The National Center on Performance Incentives is a research and development center funded in part by the United States Department of Education's Institute of Education Sciences (R305A06034). The views expressed in this research brief do not necessarily reflect those of the sponsoring agencies.

The National Center on Performance Incentives is led by Peabody College of Vanderbilt University in partnership with the RAND Corporation and the University of Missouri-Columbia.