Journal of Economic Perspectives—Volume 24, Number 3—Summer 2010—Pages 119–132

Aiming for Efficiency Rather than Proficiency

Derek Neal

The No Child Left Behind Act (NCLB) of 2001 required states to develop accountability systems that monitor the progress of all schools toward the goal of all children being proficient in reading and math by 2014. The act requires that states develop ways of measuring whether or not schools are making Adequate Yearly Progress (AYP) toward this goal, and it also requires that schools face sanctions and possible reorganization if they persistently fail to meet AYP targets.

However, the No Child Left Behind legislation is conceptually flawed for at least three reasons that are linked to its focus on proficiency. First, because the meaning of proficiency is vague, reports concerning the proficiency of students are obvious targets for political manipulation. These concerns are heightened by the fact that NCLB delegates to the states the tasks of defining proficiency standards and measuring student performance relative to these standards.

Second, universal proficiency is not an appropriate policy goal for an education system. Because each student has a fixed amount of time and energy, the marginal returns to investments in any one student diminish as investment levels increase. Thus, it may be either extremely costly or impossible to bring those students who are least prepared for school all the way up to a proficiency standard that is demanding for most students. Under common assumptions about social objectives, it is clear that efficient education policy should seek to mitigate skill differences created by differences in home environment and student preparation. However, given these

■ *Derek Neal is a Professor in the Department of Economics and the Committee on Education at the University of Chicago, Chicago, Illinois. He is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. His e-mail address is ⟨d-neal@uchicago.edu⟩.*

doi=10.1257/jep.24.3.119

same assumptions, the presence of diminishing returns implies that it is wasteful to set the exact same proficiency standard for all students regardless of their baseline skills. As a practical matter, some schools have responded to NCLB's demand that all students reach a demanding proficiency level by ignoring students who have no chance of reaching such levels without extraordinarily costly interventions, and some states have responded by compromising the meaning of proficiency.

Third, given any reasonable definition of proficiency, the threats within the No Child Left Behind legislation of severe sanctions for failure to reach 100 percent proficiency by 2014 are not credible. Very few schools will ever literally achieve 100 percent proficiency. Therefore, the text of NCLB creates confusion and uncertainty concerning how the law will actually be enforced in the future. This uncertainty may weaken the performance incentives contained in NCLB as we move closer to 2014.

In sum, NCLB offers an accountability system that combines vague, manipulable, and likely inefficient proficiency standards with threatened but noncredible sanctions for failure to meet these standards.

Even if it were socially desirable for the federal government to guarantee that 100 percent of children meet proficiency standards in math and reading, the federal government does not have the capacity to achieve this goal. The federal government provides only a small portion of the funding for public schools and has limited control over the resources that states and local districts allocate to education. In addition, the federal government has even less influence over the decisions of parents concerning how much they invest in the cognitive and emotional development of their children. The most the federal government can hope to do is to design accountability systems that require schools receiving federal funds to demonstrate that they are using the federal, state, and local funds they receive efficiently. I will argue that the goal of accountability systems should be to organize competitions among schools that serve similar students and let the outcomes of these contests dictate which schools receive government funding and to what extent. Well-organized competition among schools is the best vehicle for making sure that schools use public funds efficiently.

If education officials pursue the paradigm of requiring public schools to compete with each other for the privilege of receiving public support, they must develop relative performance measures that assess the outcomes of these contests while making reasonable allowance for differences in student populations served by public schools. Here, I discuss two different types of relative performance measures. First, value-added models produce measures of the contribution of schools or teachers to the achievement growth of their students. Most applications of value-added models produce universal rankings of educator quality because these models typically assume that quality is a fixed characteristic of the educators serving in a particular school.

However, relative performance measures in education should take seriously the possibility that some teachers may be particularly effective with some types of students but not with all students. Thus, I will also discuss a method for deriving context-specific measures of school performance. A percentile performance index tells public officials how often the students in a particular school or classroom perform better than students in other schools who began the year in similar circumstances with respect to their prior achievements, the compositions of their classmates, and their family backgrounds.[1] This index of relative performance employs only the ordinal content of assessment results but still provides the information policymakers need to make preliminary judgments concerning when to reorganize a given school and give a new staff the opportunity to prove they can do better.

The first concern of policymakers should not be whether students are reaching proficiency targets. The first order of business is determining whether students are learning as much as possible given the resources allocated to schools. (Of course, socially optimal policy also requires that authorities allocate efficient levels of overall spending for schools, but given any level of total spending, efficiency requires that resources not be wasted.) The performance frontier that corresponds to "as much as possible" is not a static concept, and it is not divorced from the family and community environments of students. Rather, relevant performance frontiers for schools are specific to the circumstances in which schools operate, and they evolve over time with developments in pedagogy and school management practices. Contests among schools that serve similar students would provide information concerning which schools are operating at these frontiers and which schools are lagging behind. Further, by focusing on the design of competitions rather than the specification of standards, policymakers may come to realize that expansions of parental choice through charter schools and vouchers are complements to, rather than substitutes for, assessment-based accountability.

Problems with Proficiency

On its surface, the No Child Left Behind legislation links the receipt of federal funds to the performance of students. Thus, it appears to be an effort on the part of the federal government to provide incentives for states and local districts to improve their performance. The proponents of NCLB appear to be motivated by the belief that state education agencies and local school districts do not always monitor school performance in an effective manner. Given this motivation, one must ask why the NCLB legislation delegates to the states the tasks of defining performance standards and measuring performance relative to those standards.

---

[1] In Barlevy and Neal (2009), my coauthor and I propose this index as part of an incentive scheme. We consider a setting in which teacher quality is known by a planner, and teachers have been assigned to classrooms optimally, but teacher effort cannot be observed. We then derive an incentive system called Pay for Percentile that induces all teachers to provide efficient levels of instruction to all of their students. Although we developed the percentile performance index for use in a specific incentive scheme designed for use in a setting with equally talented teachers, it may also prove useful as a device for gathering information about differences in teacher talent. Briggs and Betebenner (2009) propose econometric techniques for estimating performance indices that are quite similar to our measure.

If the education officials in a state believe that it is in their political interest to improve schools, they can adopt and implement their own state-level accountability system without intervention from the federal government. On the other hand, if these officials do not believe it is in their interest to hold local schools accountable, then allowing these officials to define and measure their own proficiency simply offers them the opportunity to escape the accountability that the No Child Left Behind legislation seeks to create. My wife and I periodically require our children to clean their rooms, but we do not delegate to them the tasks of defining standards for cleanliness nor of conducting the inspections of their rooms that determine whether these standards have been met.

There are many ways to compromise the meaning of a proficiency standard, and even more ways to manipulate whether progress is being made toward meeting that standard. For example, states can set proficiency standards with a less-than-demanding cutoff score so that almost all students will meet the standard, and a recent literature suggests that some states are pursuing this approach (for example, Banderia del Mello, Blankenship, and McLaughlin, 2009; Cronin, Dahlin, Adkins, and Kingsury, 2007; Dee and Jacob, 2009).

States can also create an illusion of progress in their proficiency counts by making assessments easier over time without rescaling them, or by introducing new assessments and falsely equating the scale of the new assessment to the scale of the prior assessment. As an example of this potential problem, Table 1 presents numbers directly from the 2003, 2006, and 2009 Illinois State Report Cards available at ⟨http://webprod.isbe.net/ereportcard/publicsite/getsearchcriteria.aspx⟩. The numbers describe statewide proficiency rates in math for third, fifth, and eighth graders in Illinois for 2002, 2005, 2006, and 2009. In 2002 through 2005, Illinois tested only third, fifth, and eighth graders under NCLB. In 2006, when Illinois began testing fourth, sixth, and seventh grades as well, the state introduced a new series of assessments for third, fifth, and eighth grades. The state then conducted studies that set the cutoff scores for proficiency on the new assessment scale at levels that were allegedly equivalent to the original proficiency standards.

The last row of Table 1 presents the fraction of the total increases in third-, fifth-, and eighth-grade proficiency rates in the State of Illinois during the 2002 to 2009 period that occurred in the one year, between 2005 and 2006, when the new assessment scale was introduced. These fractions are .58, .28, and .82 respectively. I know of no change in state policies governing teacher hiring, training, or classroom practice that could account for this result, and similar jumps in proficiency rates do not occur in any other year.

Do the results in Table 1 constitute definitive proof that Illinois officials manipulated the 2006 studies and the new assessment scale to create an illusion of progress toward greater proficiency? No, and such a claim is not the point. The point is that the reliance of No Child Left Behind on state-specific proficiency standards offers states opportunities to manipulate these standards in ways that

Table 1
NCLB Math Proficiency Rates for Illinois

	3rd grade	*5th grade*	*8th grade*
2002	74.2	62.8	52.5
2005	79.2	73.1	54.3
2006	85.6	78.6	78.2
2009	85.2	82.4	81.7
Fraction of total increase in scores over 2002–2009 that occurred between 2005 and 2006	**0.58**	**0.28**	**0.82**

Source: Data from the 2003, 2006, and 2009 Illinois State Report Cards available at ⟨http://webprod.isbe.net/ereportcard/publicsite/getsearchcriteria.aspx⟩.
Note: The last row of Table 1 presents the fraction of the total increases in third-, fifth-, and eighth-grade proficiency rates in the State of Illinois during the 2002 to 2009 period that occurred in the one year, between 2005 and 2006, when the new assessment scale was introduced.

are costly for federal officials to detect. The methods used to equate psychometric scales are quite complex, and the application of these methods requires elements of both art and science. Any plan that sought to monitor the integrity of proficiency standards in all states, subjects, and grades would be complex and costly.

Aiming at the Wrong Target

Many advocates of No Child Left Behind may respond that the legislation should simply be amended to include national standards and a national system of assessments that measures proficiency relative to these national standards. However, before heading down this road, one must consider whether it makes sense to build accountability systems around the concept of proficiency in the first place. I contend that under quite general conditions, it does not.

Consider an education authority that is charged with educating a number of students. Each student will start the school year at one skill level and end it at another skill level. The school has limited resources for teaching, and each student has a fixed amount of time and energy. Now, assume that the authority seeks to maximize a social welfare function that satisfies the following two conditions: First, any improvement in the skill of any child is a socially good thing, assuming that the skill levels of others are not harmed, and second, skill improvements achieved by less-skilled children are more valuable to society than comparable improvements enjoyed by children with higher baseline skills.

In this setting, the education authority will allocate education resources among students so that the returns to tutoring are equalized *at the margin* over all students. This policy will narrow skill differences in the population because the authority is guided by a social welfare function that places higher weight on skill improvements among less-skilled children. However, this authority will never allocate resources in a manner that brings students who began the year at quite different baseline skill levels exactly up to some common end-of-year proficiency level.

Consider two students who begin the year at different baseline skill levels. Assume that it will require five hours per day of after-school tutoring to bring the student with lower baseline skills all the way to proficiency by the end of the year. However, it will take only 30 minutes per day to bring the more-able student up to proficiency. It is never optimal to simply allocate five hours of extra tutoring per day to the first student and 30 minutes per day to the second. During the last 30 minutes of the first student's five-hour tutoring sessions, that student will learn relatively little, due to the fatigue induced by a twelve-hour school day. Given a fixed education budget, society will always be better off if the student with higher baseline scores receives more than 30 minutes of tutoring and finishes the year beyond the proficiency standard while the less able student receives less than five hours of tutoring per day and falls short of the proficiency standard.[2]

Accountability systems built around proficiency standards will not generate efficient outcomes because the law of diminishing returns guarantees that it will always be inefficient to invest the resources required to bring the students with the lowest baseline scores all the way up to a common proficiency standard. Society can always do better by investing slightly less in those who begin with the lowest skill levels and slightly more in those whose baseline skills placed them near proficiency beforehand. Further, because the return to additional instruction may be zero at some point due to limits on the time and energy of children, it will not even be technologically feasible to bring some children all the way up to some proficiency standards.

A significant literature documents what appear to be less-than-desirable responses by educators to the introduction of accountability systems built around proficiency standards, and I will turn to this literature shortly. However, the most important problem with proficiency-based systems is not that educators game them. The fundamental problem with proficiency standards is that they point educators toward the wrong goal. The education authority in the scenario above has the power to issue instructions that allocate precise amounts of instruction to individual students, but if one assumes that this powerful authority behaves optimally, it is certain that the authority will not use the goal of universal proficiency as a guide for allocating instructional resources.

Some may counter that, by linking sanctions and proficiency rates, No Child Left Behind at least compels schools to invest more heavily in their most disadvantaged students and therefore it will minimize the number of students who reach adulthood without the basic skills required for success in modern economies. However, the evidence suggests that this claim is unwarranted.

In Neal and Schanzenbach (2010), Diane Schanzenbach and I examine test scores in Chicago following the introduction of No Child Left Behind and also the introduction of a citywide accountability system in 1996. Our study is unique because the tests used in the first years of both accountability systems had been in place for some time before these systems were announced. Thus, we are able to estimate how achievement evolves on these assessments when they are taken under low stakes in two different grades. Then, by examining the outcomes of students who took assessments both in the first years of an accountability program and several years earlier when the assessment involved no stakes, we can measure the change in test scores induced by the accountability systems at different points in the baseline achievement distribution. We consistently find that proficiency-based accountability systems offer no benefits for students who have no reasonable chance of becoming proficient in the near term, and students at the very bottom of the baseline skill distribution may well have been harmed by the introduction of No Child Left Behind.

---

[2] For a basic mathematical model that derives these results more formally, see the online Appendix available with this paper at ⟨http://www.e-jep.org⟩.

A significant ethnographic literature on "education triage" details the strategies that educators employ to shift resources away from particular students in response to high-stakes testing systems built around cutoff scores. The logic of diminishing returns implies that it may be very costly or impossible to bring some students all the way up to a demanding proficiency standard, and it is not surprising that when education authorities measure school performance using proficiency counts, some schools ignore these students as well as their gifted classmates who are already proficient.[3]

Federal education officials have recently approved several waivers that allow states to measure the performance of their schools based on the growth of students toward proficiency rather than proficiency per se. This approach may mitigate the tendency for teachers to respond to No Child Left Behind by ignoring their least able students. However, these waivers leave states with systems that are still fundamentally flawed because the growth model plans adopted by states must, as explained at ⟨http://www.ed.gov/admins/lead/account/growthmodel/proficiency.pdf⟩, "Ensure that all students are proficient by 2014, and set annual goals to ensure that the achievement gap is closing for all groups of students." In the end, NCLB still requires that all students be held to the same proficiency standards, and this is the wrong goal for education policy. Further, it is easy to see that threats to sanction schools for not achieving this goal are not credible.

Credibility

I will not explore all the gory details of the Adequate Yearly Progress calculations that determine whether schools are making progress toward the goal of

---

[3] See Gillbourn and Youdell (2000), Booher-Jennings (2005), and White and Rosenbaum (2007). Reback (2008) also provides evidence using large data sets on individual assessment results from Texas. Since he does not have access to assessments taken under low stakes, he uses the details of the Texas accountability formulas to isolate children who offer high expected returns from extra attention and tutoring because their progress is given large weight in the calculations of their schools' performance ratings.

100 percent proficiency. However, I note three features of the law that will combine to raise concerns about the credibility of a 100 percent proficiency requirement as 2014 approaches.

First, all state plans must include a transition path toward requiring that all students be proficient by 2014, and along the way, schools are judged according to whether they are making Adequate Yearly Progress with respect to the growth of proficiency rates in many defined subgroups of students. Second, the provisions of No Child Left Behind that specify how schools will be sanctioned for failing to meet Adequate Yearly Progress deal with the performance of each school in isolation. The threat of sanctions is not influenced by whether a particular school is the only school in its state failing to make Adequate Yearly Progress, or whether it is one of several thousand such schools. Finally, the sanctions for failure to meet Adequate Yearly Progress are the same regardless of the distance between a failing school's proficiency rate and its target.[4]

The following hypothetical scenario demonstrates how these features may soon interact in ways that create a credibility problem. Consider a hypothetical school in a disadvantaged neighborhood in Chicago in 2017. Assume that 95 percent of the student population in this school is eligible for free or reduced-price lunches and that 90 percent of the students are proficient in math and reading. However, with No Child Left Behind requiring 100 percent proficiency by 2014, also assume that this school has failed to meet its proficiency rate targets for five straight years.

One expects that, among the dozens of schools in the Chicago area that serve populations of severely disadvantaged students, 90 percent will be one of the highest proficiency rates reported in 2017. Thus, given this scenario, would federal officials force the state to close our hypothetical school or to reorganize it? Would federal officials force the state to transfer resources away from this school to private agencies that provide supplementary services? Could it ever make sense to take resources away from a school that performs better with its own mix of students than the vast majority of schools that serve similar students? Finally, if hundreds of schools in a large city were to face the threat of reorganization at the same time, wouldn't the leaders of these schools know that it is simply not logistically possible to sanction all of these schools at once?

The goal of accountability systems should be to create incentives for schools to teach students as much as possible given their pre-school preparation, the support they receive from their families, and the resources available to their schools. For some students, "as much as possible" may fall short of a particular notion of proficiency, and for these students, the text of No Child Left Behind prescribes a chaotic cycle of school sanctions and reorganizations. Because such chaos will harm both children and educators, I predict that the most severe sanctions spelled out in NCLB will never be enforced, and I also conjecture that educators understand this. In 2001, the proponents of NCLB may have believed that spelling out severe penalties for all schools that do not achieve 100 percent proficiency by 2014 was politically valuable. However, as we near 2014, this requirement may make the political costs of enforcing the law unbearable, and thus render NCLB impotent as an accountability system.

---

[4] The Safe Harbor provisions in No Child Left Behind allow some schools to meet Adequate Yearly Progress by making sufficient progress toward their target, but if the law is enforced as written, the 100 percent target will become relevant for all schools at some point in the future.

Focus on Relative Performance

Federal accountability policies should shift their attention from proficiency to efficiency. There is no reason to believe that the primary goal of the education system should be defined by an effort to make literally every child proficient based on a one-size-fits-all notion of proficiency. Further, the federal government does not have the capacity to compel states and localities to devote the resources required to make all children proficient. However, federal accountability plans based on competition, rather than proficiency standards, have the potential to encourage state and local governments to use the resources they devote to education more efficiently.

Policymakers do not possess the information required to set efficient achievement targets for every student that vary appropriately with individual differences in background and prior achievement. In addition, even the achievement levels we should expect from various types of students are constantly evolving, because teaching techniques and our understanding of the ways that different children learn evolve over time. Thus, the best way for policymakers to promote efficiency is to build accountability systems around organized competitions among schools that reveal how achievement frontiers are evolving over time for different types of students. The measures of school performance that determine the winners and losers in such competitions are not measures of performance relative to some set of proficiency standards but measures of performance relative to other schools that serve similar students.

Among educational statisticians, value-added models are the most commonly advocated tool for creating relative performance measures. These models work as follows: Use a vector of baseline student characteristics, $\mathbf{Z}$, to create expected student test scores at the end of a school year given each student's characteristics. Then, at the end of the year, measure the average deviation of actual student scores from these predicted scores among all the students in a given school. At the school level, these average deviations provide a one-dimensional performance measure that serves as a universal quality ranking over all schools.

While this approach seems appealing in many respects, problems arise in practice. For example, if there is little or no overlap in the baseline achievement distributions for schools A and B, the estimated difference in overall quality implied by a value-added method or any other method of creating universal rankings is to a large extent the creation of the researchers, and value-added methods may conceal the fact that the data provide no direct evidence concerning how two schools perform with respect to a set of comparable students.

In addition, the universal quality rankings generated by a value-added method may provide answers to irrelevant questions. Assume that school A is a neighborhood school that serves disadvantaged students and that school B is a magnet school with selective admission. Assume that although most students would expect higher achievement in the magnet school B than in the neighborhood school A, the students in school A would not perform better in the magnet school. In this setting, value-added methods may produce results that imply that school B is better than school A, but why should this matter to the students in school A? School B is not better for them.

An alternative means of creating relative performance measures for schools is the percentile performance index described in Barlevy and Neal (2009). Assume that for each student in each grade, policymakers can estimate a conditional distribution of end-of-year assessment scores, given any vector of baseline personal characteristics $\mathbf{Z}$ that shape the costs of bringing the student up to various levels of achievement. When forming an estimator of this distribution, policymakers may employ measures of past achievement, family resources, past achievement by peers, and other resource measures as elements of $\mathbf{Z}$.[5]

Given the expected distribution of end-of-year scores, together with individual results on an end-of-year assessment, education officials can assign each student a percentile score that denotes the percentile score for that student among all students in the school system that share the same vector of baseline characteristics $\mathbf{Z}$. The average value of this percentile score in a particular subject over all the students in a given grade is the percentile performance index for that school with respect to a given grade–subject combination: for example, fifth-grade math. One can calculate a schoolwide percentile performance index by forming a weighted average of the index for each individual subject and grade. If a school has a percentile performance index of X, it implies that the probability a given student in this school will perform better than a randomly selected student with similar baseline characteristics who attends a different school is X. Thus, this index is an estimated winning percentage for a particular school that describes the success of its students in implicit competitions against students in other schools who share similar baseline achievement levels and family background characteristics.

The optimal way to employ a percentile performance index when making decisions concerning which schools should be declared ineligible for government funding remains a matter for future research. However, if a school has a percentile performance index around, say, 0.10 for several years, we know that other educators who are working with students elsewhere from similar backgrounds are almost always getting better results. This suggests that other educators may be able to export pedagogical approaches to this school and improve student achievement without a need for additional resources. (Of course, this assumes that all schools in a district or state receive comparable resources for classroom activities. If this is not the case, percentile performance index measures should compare schools that have access to similar classroom instruction budgets.)

---

[5] I will not comment here on the different methods that may be employed to produce estimators of the distribution of scores given $\mathbf{Z}$ other than to note that estimating conditional distributions of this type will require large data sets and may require additional state-wide integration of data systems. But for those interested in detail about how this might be done, Briggs and Betebenner (2009) develop an estimator along these lines by employing a series of quantile regression models and implement the estimator using data from Colorado schools.

Percentile performance indices do not allow policymakers to form a universal ranking of school quality because these percentiles are ordinal comparisons among students who were comparable at the beginning of the academic year. The percentile performance index approach summarizes how well a school performed against its implicit competitors, but because the identity of these competitors varies according to the baseline characteristics of the students in a school, this index is not an attempt to measure school quality in a consistent manner for all schools. Nonetheless, this approach still allows policymakers to identify schools that are clearly inefficient. Schools with persistently low percentile performance indices have students who consistently perform worse than their peers in other schools, which is prima facie evidence that these schools are not using their resources efficiently.

Further, because the percentile performance index approach is based on the ordering of scores as measured by percentile, its results are invariant to the scales that testing agencies choose for reporting assessment results (as long as different possible scales preserve order). In comparison, value-added measures of school performance do depend on scale, and as a result, these measures serve as accurate measures of relative quality only if one is willing to treat the units of a particular test score scale as a social welfare index.

Essentially, value-added models are based on an underlying assumption that test scores or some transformation of test scores are reported in units such that, whenever a student gains one point, this gain is of equal value to society regardless of the student's prior level of skill. However, I know no reason that society should assume that a student whose math score moves from, say, 120 to 122 creates the same gain for society as another student whose score moves from 220 to 222. Value-added methods weight all deviations from predicted score gains equally. But, it seems reasonable to suspect that, over certain regions of a given scale, score gains may be both more valuable for students to achieve and more costly for teachers to foster, and in this case, the average test score gains that value-added methods associate with school quality are not valid measures of school performance.

Conclusion

The No Child Left Behind law is flawed for many reasons, but the most important is that it is built around proficiency targets. Proficiency rates are not useful metrics of school performance because universal proficiency is not a socially efficient goal for principals and teachers. Further, the variation in proficiency rates among schools reflects, in large part, interschool differences in student background characteristics. The designers of accountability systems must move away from systems designed around a one-size-fits-all standard and begin designing systems that organize and promote competition among schools.

This emphasis on competition between schools naturally raises the question of a possible role for charter schools, greater choice between public schools, voucher systems, and the like. For example, the optimal system may involve providing parents with the information produced by student assessments and perhaps other information like school inspections and then letting parents direct resources to the schools they prefer. More work remains concerning the design of these assessments and inspections as well as the design of the system used to communicate results to parents. School choice and voucher systems also require specifying rules concerning whether schools can practice selective admission, how the generosity of vouchers varies with family income, and whether schools should be allowed to charge tuition in excess of the voucher level. For example, MacLeod and Urquiola (2009) offer an analysis of how rules concerning the use of selective admission in voucher schools affects the efficiency of voucher systems.

Nonetheless, with these reservations duly noted, it is important to recognize that while academic tests may do a reasonable job of capturing the growth of cognitive skills, parents have access to valuable information about the social and emotional development of their children. A growing literature in economics and psychology (for example, Cunha and Heckman, 2008) stresses the importance of noncognitive skills that children acquire before heading into the worlds of higher education and work. Expansions of parental choice in whatever guise could allow government to acquire an army of educational performance monitors. Thus, the best educational accountability systems might treat test-based accountability as part of a larger accountability system in which the amount of public funding available to a given school is determined by the assessments results of students in the school, the number of parents who choose the school, and results of inspections that assess the safety of the school's physical environment. In Neal (2009), I discuss the design of such incentive systems for publicly funded schools.

To date, economists working in education research have conducted careful and often sophisticated empirical evaluations of existing accountability and incentive systems. However, the economics literature on the optimal design of these systems is much smaller, even though economists, at least relative to other social scientists who work on education policy, should have some expertise as designers of incentive systems. Given the flawed designs of the No Child Left Behind legislation and many similar state accountability systems, it may be time for economists working in education research to alter their focus.

■ *I thank Kevin Lang, Gadi Barlevy, and Diane Schanzenbach for helpful comments. I thank the Searle Freedom Trust for generous research support as well as Lindy and Michael Keiser for support through a gift to the University of Chicago's Committee on Education.*

References

Banderia del Mello, Victor, Charles Blankenship, and Don McLaughlin. 2009. *Mapping State Proficiency Standards onto NAEP Scales: 2005–2007.* Research and Development Report (NCES 2010–456). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Barlevy, Gadi, and Derek Neal. 2009. "Pay for Percentile." Federal Reserve Bank of Chicago Working Paper no. 2009-09, November.

Booher-Jennings, Jennifer. 2005. "Below the Bubble: 'Educational Triage' and the Texas Accountability System." *American Educational Research Journal,* 42(2): 231–68.

Briggs, Derek, and Damian Betebenner. 2009. "Is Growth in Student Achievement Scale Dependent." Available at: http://www.colorado.edu/education/faculty/derekbriggs/publications.html.

Cronin, John, Michael, Dahlin, Deborah Adkins, and G. Gage Kingsbury. 2007. "The Proficiency Illusion." Unpublished paper. Thomas B. Fordham Institute, October.

Cunha, Flavio, and James J. Heckman. 2008. "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Journal of Human Resources,* 43(4): 739–80.

Dee, Thomas S., and Brian Jacob. 2009. "The Impact of No Child Left Behind on Student Achievement." NBER Working Paper 15531.

Gillborn, David, and Deborah Youdell. 2000. *Rationing Education: Policy, Practice, Reform, and Equity.* Philadelphia, PA: Open University Press.

Illinois State Board of Education. 2003, 2006, and 2009. Illinois State Report Card. (You may request them at ⟨http://webprod.isbe.net/ereportcard/publicsite/getsearchcriteria.aspx⟩.)

MacLeod, Bentley, and Migual Urquiola. 2009. "Anti-Lemons: School Reputation and Educational Quality." NBER Working Paper 15112.

Neal, Derek. 2009. "Designing Incentive Systems for Educators." Chap. 7 in *Performance Incentives: Their Growing Impact on American K-12 Education*, ed. Matthew G. Springer. Brookings Institution.

Neal, Derek, and Diane Whitmore Schanzenbach. 2010. "Left Behind By Design: Proficiency Counts and Test-Based Accountability." *Review of Economics and Statistics,* 92(2): 263–83.

Reback, Randall. 2008. "Teaching to the Rating: School Accountability and the Distribution of Student Achievement." *Journal of Public Economics,* 92(5–6): 1394–1415.

White, Katie Weits, and James E. Rosenbaum. 2007. "Inside the Blackbox of Accountability: How High Stakes Accountability Alters School Culture and the Classification and Treatment of Students and Teachers." A section in *No Child Left Behind and the Reduction of the Achievement Gap: Sociological Perspectives on the Federal Education Policy.* ed. A. R. Sadvonik, J. A. O'Day, G. W. Bohrnstedt, and K. M. Borman. Routledge.

This article has been cited by:

1. Derek NealThe Design of Performance Pay in Education **4**, 495-550. [CrossRef]