United States Lessons about School Accountability

Eric A. Hanushek *

The United States has launched a new experiment designed to improve its schools. The most publicized portion of this is the current federal educational policy to expand school accountability based on measured student test performance. Although many states had already installed accountability systems by 2000, a central campaign theme of George W. Bush was to expand this to all states, something that became a reality with the No Child Left Behind Act of 2001 (NCLB). The landmark NCLB codified a developing policy view that standards, testing and accountability were the path to improved performance. This discussion provides evidence on the expected effects of NCLB not only on student performance but also on other potential outcomes.
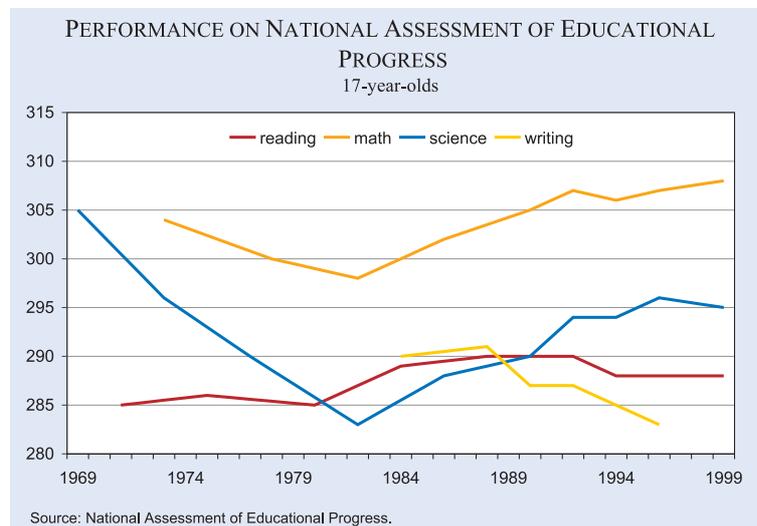
Accountability has been a catchword in education for decades. While it seems natural to measure outcomes and hold schools responsible for them, the mechanics of how to do that appropriately are complicated. Considerable controversy accompanies accountability in schools. Parents, teachers, policy makers, and the American public frequently enter into debate about various elements and uses of accountability systems. These debates are motivated by different underlying views about how best to improve the education of our youth as well as by self-interested reactions.

Understanding the dynamics of the United State education system sheds light on the current thrust towards accountability and the issues facing today's policy makers. In simplest terms, student performance has stagnated while costs have steadily increased. These simple facts have led to the realization that just providing more resources within the current structure is unlikely to be effective. Nor does adding further regulation offer much promise.

This stagnation is illustrated by the results of the National Assessment of Educational Progress (NAEP), which regularly tests students across the country in different subject areas. The tests, which have been conducted over the past three decades, start with a random sample of students from different grade levels. A summary of the performance of 17-year-olds over time is provided in the Figure. The Figure tracks average scores in reading, math, science and writing. The story is one of flat achievement. Reading and math scores are slightly higher at the end of three decades, while science and writing appear to have noticeably declined.

Level performance would not be a matter of serious concern except for two important additional trends. First, it parallels mediocre performance on the international level, where the United States has performed at or below average since the

PERFORMANCE ON NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS
17-year-olds

Source: National Assessment of Educational Progress.

* Eric A. Hanushek is Senior Fellow at the Hoover Institution, Stanford University.

Public school resources in the United States 1960–2000

	1960	1980	2000
Pupil-teacher ratio	25.8	18.7	16.0
Percentage of teachers with master's degree or more	23.5	49.6	56.2[a)]
Median years teacher experience	11	12	15[a)]
Real expenditure/pupil (in US$)	2,235	5,124	7,591
[a)] Data for 1996.			

1960s.[1] Second, the US performance has not been for want of trying. As the Table shows, school resources have been increased over the same period of time. Real spending per student more than tripled between 1960 and 2000. This increase in resources was accomplished in the way typically called for by reformers and policy makers: by significantly reducing pupil-teacher ratios, by increasing the training of teachers and by developing a more experienced teaching force.

The dominant approach to policy making over much of this period has been regulation of education inputs and processes along with providing resources for specific school programs. This approach has been especially appealing to legislatures, because it is easy to set resource policy. But, as shown in the aggregate data, increased resources have not improved performance. Moreover, these overall impressions have been reinforced by similar findings of analyses of performance across classrooms and schools (Hanushek 2003).

This lack of improved performance has brought attention to alternative means of effecting change in schools. This attention has been manifested in a variety of forms, but a common theme has been the regulation of outcomes rather than the more traditional regulation of process and inputs. Previous efforts were based on providing or prescribing specific inputs (such as reduced class size in specific circumstances) and hoping that these led to improved student performance. Often, however, these decisions were based on little information that would indicate high probabilities of success. The new regulatory frameworks tend to emphasize objective outcomes, while letting schools decide how they would meet demands for achievement. The underlying

idea is that public monitoring and reporting of student outcomes, perhaps coupled with consequences for schools, teachers or students will be more successful than previous input-based policies.

Design of accountability systems

Each state in the US has developed its own student accountability system, and, while NCLB provides some guidelines, the states remain pre-eminent in developing school policy and accountability. NCLB did, however, dramatically change the focus of schools by declaring that all students would achieve some measure of proficiency within a dozen years.

The basic skeleton of accountability systems involves content standards, measurement, consequences and reporting. While states differ in significant ways, a general description of the structure of these systems is useful in comparing actual plans and how their elements interact.

Content standards

Content standards typically present the details of what is expected in each subject and grade. They create boundaries or domains for attention. The typical student outcome standards delineate a body of material that has been designated by an authoritative body to represent a minimum acceptable set of knowledge to be mastered by the student. Although apparently straightforward, the creation of precise standards has been fraught with difficulty. Tension exists between the need for a representative set of elements and the need for the elements to be testable (discussed below).

Measurement

The biggest controversy in accountability probably surrounds how compliance with standards should be measured. Proving that the standards have been met requires some sort of measurement. Assessing compliance requires several decisions: who to measure; what approach to use; how to create valid indices; and, frequently, where to set the critical value or cut-point for meeting the standard.

The centerpiece of current state accountability systems is the testing of student performance. This performance is then aggregated to, say, the school

[1] At least in recent years, these results do not reflect international differences in selectivity of schooling or test taking but instead appear to reflect more fundamental forces. A summary of the performance of countries across the tests along with references to the basic data can be found in Hanushek and Kimko (2000).

or district level, and some summary of the test scores is made public.

Deriving composite measures

While most of the public attention has gone to the development of standards and how to measure compliance with them, the use of resulting data, particularly when there are multiple objectives, is equally important. The goal of an accountability system is improving student performance, but performance is the outcome of a variety of factors: student ability and effort, parental inputs, teacher inputs, and school programs and resources. Even with accurate and reliable data on student performance, the outcome statistics produced must reflect the actions of the actors if they are to enter appropriately into performance incentives.

The issue of disentangling underlying elements of performance is most frequently raised in assessing the performance of teachers and schools. If we take accountability down to each schools and teachers, it is common sense that none should be held responsible for bad performance by others. For example, if a teacher starts with low-performing students but does a terrific job of improving their performance, she should not be penalized if the resulting performance level is still lower than, say, the national average. Similarly, a teacher starting with a high-performing group should get credit for her job in improving them but not for their initial preparation. The implication is that any measurement of teacher quality should focus on the teacher's addition, or value-added, to student learning – and this requires adjusting the measurement of student performance according to the initial preparation of students.

The best way to separate the different factors that influence student performance is not obvious. A variety of approaches has been proposed and experimented with in the states. The most common starting measure – applied in virtually every existing accountability system – is the average of all student test scores for a district or a school. This aggregate summary, however, mixes all sources of performance. Extensions that have been proposed and used in different places include:

• Annual change in school average score over time
• Average of the mean individual gains in scores

• Average scores of a school relative to state average scores for students of similar background
• Regression adjusted scores to remove individual background differences.

These measures also highlight a fundamental tension between the incentives that are created by the way a given accountability system is structured and the overall performance goals they are supposed to promote. For example, for many uses it may be desirable to pinpoint the value added by each school, but even a high value added school may start with students sufficiently ill-prepared that the school does not bring them up to the desired levels of student performance. Looked at from the viewpoint of enforcing high standards of student performance, this school might be judged as falling short – while from the incentive side, this school would deserve praise. This apparently simple issue illustrates the difficulties of using student performance data simultaneously for multiple goals. A common approach is for states to create incentives involving a combination of the level of score and the school change in score over time (such as seen in school reward systems in North Carolina and California).

Uses and consequences

Goals, standards, and measurements create an accountability system. But in most states, accountability systems have multiple objectives – including creating a measuring rod for outcomes, improving school instruction, creating incentives and creating rewards/punishments for performance.

The standards and accountability movement strives to induce alignment between standards, teaching and student performance. In contrast to a regulatory approach, the underlying philosophy of accountability is letting the actors maintain control of a process whose outcomes are scrutinized. Consequences – both positive and negative – are the fulcrum that gives leverage to the other players in the education system. If schools or students do not expect any decisive actions as a result of their performance, there is little to motivate attention to the outcomes they produce.

No child left behind

The No Child Left Behind program took the state systems that were developed and layered on new aspects that related largely to consequences if

schools did not show improvement in student scores. Specifically, NCLB required states to develop a plan for ensuring that their students will eventually be proficient in the basic subjects. A key element is tracking whether schools are meeting "adequate yearly progress". This is a measure of whether the state and the school are following on an acceptable trajectory.[2] If schools are not performing well by these standards, they face increasing requirements to support the students in the schools with more choice of schools, with supplemental services and ultimately with the ability to leave a public school completely and to find a better school. The school itself may even, at some point, be dissolved.

The key element added by NCLB is a set of sanctions against schools that perform poorly. These sanctions require, at least in principle, that the school aid the student in finding better schooling that brings the student to proficient levels. The federal program, however, is only one element. The states include their own rewards and sanctions for schools that do well or do poorly. A clear change has been the universal introduction of consequences (in the future) for schools that do not perform well as measured by student achievement.

Assessing the results

Although it is not possible to investigate the impact of NCLB directly, it is possible to assess the impact of the underlying state accountability systems that were introduced before the federal legislation. This discussion summarizes a more detailed statistical analysis found in Hanushek and Raymond (2005).

The inability to assess NCLB comes from a variety of factors. First, and most importantly, the majority of states had already instituted some sort of accountability system by the time the federal law took effect. Although only 12 states had accountability systems at the school level in 1996, 39 states did so by 2000. Thus, there is no ready comparison group that can indicate what might have happened without any law. Second, the law has many facets making it hard to isolate the effects of any single

one. Finally, the common pace of implementation of NCLB across the states eliminates any status quo alternatives for comparison.

Nonetheless, it is possible to assess how the prior state accountability systems impact student outcomes. The varied introduction of accountability across states during the past decade permits analysis of how this institution affects student learning. Moreover, because NCLB is based on individual state systems, the prior state structures are generally very closely related to their responses to the federal legislation.

The basic analytical approach builds on information about state differences in mathematics and reading performance as identified by the National Assessment of Educational Progress (NAEP). Although these tests are not the ones employed by the states in their accountability systems, they do provide a common assessment of performance across the states (something that is not available with the individual and different tests used by each state). We directly analyze how performance growth on each test between grade 4 and grade 8 is affected by whether or not a state applies a test-based accountability system.

Extensive analyses of educational production functions have been conducted, and they form the relevant background for this work. Those studies have concentrated on describing how various inputs to schools enter into the determination of student outcomes. As described elsewhere, however, these studies have not provided any consistent picture of how schools affect student performance (Hanushek 2003).

Many different state policies, regulations, and incentives – although poorly identified and measured – enter into determining student performance. State educational policy encompasses a wide range of factors including financial structure, collective bargaining rules and laws, explicit regulations on educational processes, and the like. The complications for the analyst are multiple.

The objective is to separate the impact of accountability from other possible influences on student achievement. But, if these other influences cannot be readily measured and cannot be directly controlled for in any statistical analysis, they are likely to be correlated with accountability, making accu-

---

[2] This discussion is a very simplified one that leaves out many details. In fact these provisions have been very controversial, in part because it is difficult to specify the ways of meeting the objective while being both flexible and serious about the results. As with most new systems, the exact provisions will undoubtedly evolve over time.

rate identification of the impacts of accountability on achievement impossible.

The analysis summarized here relies on three related approaches to the statistical disentanglement of the impacts of accountability. First, important components of general state factors that influence achievement – either from policy or from the character of the state population – will influence student achievement at multiple levels of schooling. Therefore, if we look at the growth of achievement over time – say, as used here, the change in student performance between fourth and eighth grades – common factors that have a constant impact on the level of achievement over the observation period will be fully captured in the early test score and thus will be implicitly controlled in the statistical analysis. Second, from the multiple observations of performance in each state over time, it is possible to estimate a common state-specific growth rate for student performance. (Technically, this amounts to extracting a state specific fixed effect in the growth equations for student scores). The easiest interpretation of this approach is that we compare growth in achievement before and after a state introduces an accountability system; that is, each state acts as its own control for comparing the effects of accountability. Thirdly, a variety of time-varying factors can be entered directly into the analysis. For example, the pattern of spending on students in each state or changes in the adult (parent) population are readily measured and can be introduced to avoid any complications that might arise from correlations with the introduction of accountability measures.

How accountability affects performance

There are straightforward findings from the statistical analysis of student achievement. We find that the introduction of accountability systems into a state tends to lead to larger achievement growth than would have occurred without accountability. The analysis, however, indicates that just reporting results has minimal impact on student performance and that the force of accountability comes from attaching consequences, such as monetary awards or takeover threats, to school performance. This finding supports the contested provisions of NCLB that impose sanctions on failing schools.

Much of the explicit interest in accountability and the federal legislation, however, focuses on low achievers. And, given the generally lower achievement of minority groups, an implicit assumption is that accountability – as revealed through mandatory disaggregation of performance for racial and ethnic groups – will simultaneously close the large racial/ethnic achievement gaps along with improving all performance. When we look specifically at the performance of subgroups, we find that Hispanic students gain most from accountability while African Americans gain least.

The finding of differential effects of accountability raises a clear policy dilemma. A prime reason for the US federal government to require each state to develop a test-based accountability system involved raising the achievement of all students, but particularly those at the bottom. It has done that, but not at the same rate across groups. We conclude from this that additional policies are needed to deal with the multiple objectives. Again, as is frequently the case, a single policy cannot effectively work for two different objectives – raising overall student performance and providing more equal outcomes across groups.

The movement toward stronger accountability in schools has also suggested to many that there would be adverse consequences – more exclusions, higher dropout rates, a narrowing of the curriculum, and the like. While some existing research supports these presumptions, it appears that the negative impacts are likely to be considerably overstated (Hanushek and Raymond 2003). Importantly, many of the adverse effects that involve "gaming" the system come from short-run incentives that are unlikely to be present over time.[3]

Although not the focus of this discussion, it is important to note that the character of currently available accountability systems is not particularly strong. This concern has two dimensions. First, the educational standards and accountability systems vary dramatically across states, and an analysis of the underlying conceptual structures suggests that different systems are expected to be associated with stronger achievement gains because of the incentives created – even though it is not possible

[3] Hanushek and Raymond (2005) provide an analysis of special education placement rates, one of the most cited areas for possible abuse, but this indicates clearly that accountability has not had an overall impact through this form of exclusions.

to distinguish clearly among alternative quality ratings in the statistical analysis.

Second, a majority of the systems concentrates on overall achievement levels (with highly variable passing scores across states). Such systems do not generally provide clear signals about the value-added of schools. Instead, they combine a variety of effects including those resulting from family background differences and neighborhood effects. As such, they cannot provide truly clear and strong incentives. Moreover, while there is a range of potential consequences incorporated in state systems, it is not possible to investigate whether some specific consequences are more powerful than others. Yet, in the face of the rather blunt incentives from existing systems, the introduction of accountability systems is associated with achievement improvements on the order of 0.2 standard deviations of student achievement. (This gain amounts to moving students from the median of the distribution to the 58th percentile). Such improvements, while not revolutionary, are notable when compared to the failure to find alternative reforms that yield such impacts on a broad basis. As accountability systems evolve, they are likely to have considerably stronger impacts if they move in the direction of more precise incentives for individual schools.

Without doubt, the achievement of our students has direct ramifications for the future well-being of society (Hanushek 2004). It should be a very high priority to ensure that all of our students do in fact gain the skills that will be needed as our economy grows and evolves. Improved school accountability offers much more promise than the failed input policies of the past, although much is still to be learned.

References

Hanushek, E. A. (2003), "The Failure of Input-based Schooling Policies", *Economic Journal* 113 (485), F64–F98.

Hanushek, E. A. (2004), "Some Simple Analytics of School Quality", *National Bureau of Economic Research Working Paper* no. 10229.

Hanushek, E. A. and D. D. Kimko (2000), "Schooling, Labor Force Quality, and the Growth of Nations", *American Economic Review* 90 (5), 1184–1208.

Hanushek, E. A. and M. E. Raymond (2003), "Lessons about the Design of State Accountability Systems", in P. E. Peterson and M. R. West, eds., *No Child Left Behind? The Politics and Practice of Accountability*, Brookings, Washington DC, 127–51.

Hanushek, E. A. and M. E. Raymond (2005), "Does School Accountability Lead to Improved Student Performance?", *Journal of Policy Analysis and Management* 24 (2).