Rewarding Teachers for Student Performance

David K. Cohen

Introduction

If professionals were perfect they would always consult carefully with clients and accurately determine the best course of action. If clients were perfect they would find the professionals they preferred, they would consult carefully to decide a course of action, and they would always work hard to do what had been decided.

But professionals and clients are not perfect. They often are unsure what to do. When they are sure they often turn out to be incorrect. And even when sure and correct they often fail to apply themselves to the solutions. There are many failures in professional practice.

The social organization of practice offers ways to manage such failures. In medicine, until recently in the United States, most doctors and patients found each other in markets. Dissatisfied patients could change physicians if they lived where practitioners were plentiful and price was not a problem. Physicians could refer or refuse cases that they found too difficult or otherwise unsuitable. Failures in practice could be remedied by choice.

But markets also fail. Many people have little or no health care, and many others who have care get only a low-grade version. They usually have little effective choice with which to remedy failures of professionalism. Some commentators argue that markets for health care create as many failures as they solve, and call for universal provision under some sort of public sponsorship or subsidy.

Elementary and secondary schools operate under one version of public sponsorship. Education is provided by state and local authorities, and all students are included without reference to their families' ability to pay. Indeed, students are compelled by law to attend school, and most schools are not only sponsored but actually operated by state agencies. Teachers and students are assigned to each other by administrative decision. There are elements of choice around the fringes—families can use residential choice to select schools, or they can pay for private education that satisfies state criteria—but most teachers and students have only modest choice at best. It is difficult for most students to find schools or teachers different from those assigned, and it is no easier for most teachers to find students different from those assigned. These limits on choice increase as family incomes decrease.

By definition one finds few market failures under this regime, but that still leaves plenty of room for failures of professionalism. Though there is some evidence that school performance has either improved or held steady as more and more disadvantaged students entered and stayed on in school since the early twentieth century, critics and reformers have increasingly argued that test scores are too low or are declining. They argue that public education has failed, and many Americans seem to share that view. They blame teachers' unwillingness or inability to offer suitable instruction, or school administrators' unwillingness or incapability to require better work from teachers, or organizational defects that inhibit solid professional accomplishment, or some combination of these conditions.

One response has been growing efforts to hold practitioners accountable for students' performance. These efforts have multiplied since the 1970s, and their popularity owes something to the simplicity of the central idea: if teachers were penalized for poor student performance or rewarded for good performance, or both, then they would have potent incentives to improve students' work and the work would improve. That was among the key ideas in state and local minimum competency testing programs that were popular in the 1970s and 1980s. It also became an important part of efforts, at about the same time, to improve education for disadvantaged children. Similar ideas turned up both in the recent congressional reauthorization of Title I of the ESEA and in the companion legislation titled Goals-2000. There also has been growing interest in broadly reorienting public schools to results produced rather than
resources applied and in using performance rewards to do so. Though commentators and policy makers disagree about how it should be done, more and more agree that some system of performance rewards is needed in public education.

One key similarity among recent performance reward schemes is their administrative character. Rather than substituting markets and consumer choice for state-administered schooling, these schemes would augment the state administration of schooling with scientific assessments of educational effectiveness and schedules of rewards and sanctions. Performance rewards are an effort to improve the operation of schools within a state-maintained framework, not to change the framework.

As policy makers, practitioners, and researchers consider such schemes, they will want to know what sorts of performance reward systems might be considered and about the grounds on which they might choose among the alternatives. I consider both matters here. I begin by delineating the key issues in the design of any performance reward scheme and then discuss alternative ways that educators and policy makers could manage these issues. I also note several issues on which research could help.

Three broad themes run through my analysis. One is that despite many efforts to institute performance reward schemes, there is relatively little reliable professional or social science knowledge about their operation and effects. We know enough to rule out simplistic and punitive schemes—though that may not be sufficient to keep them from being adopted—but not enough to inform choices among more complex and sophisticated alternatives. A second theme is that performance rewards alone cannot solve the problems of weak school performance. Though we do not know nearly enough to design the best performance reward scheme, we know that even the best scheme would require critical additional elements to improve performance in the schools in which performance is worst. Yet much of the appeal of performance rewards lies in the idea that they would be a relatively simple and cheap solution to a complex and difficult problem.

A third theme is that though performance rewards would require unprecedented clarity about measures of performance, criteria of success, and fairness, these are matters about which Americans deeply disagree. The current schooling regime inhibits clarity because schools have been the vehicle for many different hopes and purposes, have been the site of many compromises over fundamental social issues, and in the process have accumulated many layers of purpose. Advocates of performance rewards propose to sweep such murk away and replace it with clear criteria of success and standards of fairness. It seems almost sure that such schemes could not work as intended if much greater clarity were not achieved, but it also seems very likely that the more clarity we achieved, the more disputes would ensue.

**Design Issues: An Overview**

When school officials consider performance reward schemes, they will have to decide several fundamental issues. To begin with, whose performances would the scheme focus on? Many accountability systems in elementary and secondary schools focus on students' performance on achievement tests. But many U.S. universities and a few school systems focus on instructors' performance in service, research, and teaching.

No less important, who would be rewarded for the performances thus produced? In the higher education schemes just mentioned, professors are rewarded for their service, teaching, and research; students' performance never enters the picture. In contrast, most performance reward schemes in K-12 education reward or penalize teachers for students' performance, an approach that assumes teachers have considerable leverage on students, or that teachers' and students' interests are quite close, or both. Different designs are possible. Students could be rewarded for their own performance, principals or central office staff could be rewarded for students' performance, or some combination of these.

Once those issues were settled, school officials would have to decide whether individuals or groups would be rewarded or penalized. If teachers were to be rewarded, would the consequences accrue to individuals or to entire faculties? Collective sanctions are thought to encourage cooperation, whereas individual sanctions are thought to encourage competition. Treating students as individuals does seem likely to encourage competition, whereas rewarding them as classes or entire schools could promote cooperation.

Finally, would measures of performance be internal or external to the performance situations? Many merit pay systems in K-12
education sought to hold teachers accountable for students' performance on tests that were created by some external agency—most often a testing or publishing firm. In such cases, teachers were held accountable for students' work on standards of performance that they had no hand in devising. But a few merit pay schemes in K-12 education, and many in higher education, are keyed to teachers' production of evidence about their performance on internally devised performance criteria. In some such cases, teachers had a relatively free hand to design their own portfolios, whereas in others, they report their performance on locally designed criteria.

These issues suggest an enormous range of ways to interpret performance rewards, but there has been less variety in practice. Most schemes of recent vintage in K-12 schooling reward teachers for students' performance on some sort of external test. Very few schemes reward teachers for their own performance, and very few reward teachers on internal rather than external criteria. Hence, I will focus on performance rewards for teachers that are keyed to students' performance on some sort of external test or examination.

Even with these restrictions, state or local school systems that were considering performance rewards would face several additional issues, including:

- Measures of performance: What sorts of student work would be rewarded?
- Criteria of success: What sorts of performance would be rewarded and punished on whatever measures were chosen?
- Professional capability: Would school professionals know enough to respond constructively to performance rewards? If not, how could such schemes be made to work?
- New and old incentives: Performance rewards would be introduced amid many extant incentives; how might the new incentives fit with existing incentives for school performance?
- Rewards and penalties: What would be appropriate ways to reward teachers for students' performance or to penalize them for poor performance?

**Measures of Performance**

In order to reward teachers for students' performance, that performance would have to be assessed somehow. Furthermore, a reward scheme could be valid only if such assessments accurately measured the performance that educators or policy makers wanted to reward, assuming that they could specify such a thing. Hence one must begin by asking: for what sorts of student performances teachers might be held accountable? Nearly all performance reward schemes contemplate using assessments of academic achievement as the central measure of performance. Hence state or local officials would have to decide what sorts of academic assessments made most sense, but they also would have to decide whether, in light of all the things that schools try to do, any such assessment offers an adequate account of desired school outcomes. Even if we begin by considering only academic performance, assessments of school achievement differ in at least three ways that would count for performance reward schemes: desired results, feasibility, and cost.

**Desired Results**

There are large differences in the ways that various assessments measure academic performance, and adopting one sort would emphasize academic outcomes to which others would give less attention. For instance, standardized norm-referenced tests tend to focus on facts and skills. Test questions offer a menu of possible answers, only one of which is right. Many researchers now argue that such tests reward students who can memorize disconnected bits and pieces of knowledge, and using such tests in a performance reward scheme would encourage didactic instruction that focused on mindless memorization.

In contrast, authentic or performance assessments are intended to encourage students to make sense of serious practical and intellectual problems and to demonstrate how they reached an answer. Often the questions have no simple, single, right answer, and students' explanations and justifications of their work can count as much as the correctness of their answers. Many researchers argue that such assessments would encourage more complex and thoughtful performances than standardized norm-referenced tests. But the more complex and thoughtful the performances, the more likely it is that expert judges would have to be employed to read and rate the answers, an endeavor that would require educating such judges to use the same criteria to judge students' work. For without highly reliable judges' ratings of the
same performances, such assessments would be deemed unfair in comparing students’ performances.

Portfolio assessment—in which teachers or students compose collections of student work within particular domains—could draw on an even greater range of performance. Advocates argue that such assessments would encourage intellectually deeper and more adventurous classroom work because they would draw on a much greater variety of work within any domain. Assessment would not need to be limited by the time or format of common testing programs. But because portfolios and authentic assessments would be complex and more or less open-ended, students’ performances would be rated by judges rather than scored by optical scanners. It is unlikely that such judgments could be comparable across diverse performances produced by students from different classrooms, schools, and districts.

Assessments thus take very different approaches to defining the academic terrain in which they purport to measure performance. Standardized norm-referenced tests are designed to sample performance within very broad domains—chiefly reading and mathematics—but to be independent of particular approaches to curriculum and instruction. That is partly because they were not designed to be useful to instruction, but to monitor performance broadly in large populations. In contrast, criterion-referenced tests are designed to cover specific curricula, and items are selected with an eye more to topic coverage than to the distribution of responses. They are intended to be useful to instruction, for they report on how much of a given curriculum or set of objectives particular students have learned. Performance and portfolio assessments are a more recent development, and they are intended to be closely integrated with instruction. Both are seen as ways to probe students’ knowledge that enable students to more fully display their capabilities and that enable teachers to learn more fully how students think and what they know. Although performance and portfolio assessments would cover specific fields of knowledge, no strong conventions exist concerning how they should do so. Several “authentic” assessments in mathematics have been drawn up with an eye on the NCTM curriculum standards, but at least some of these have paid as much attention to extant curriculum as to any more careful consideration of the field in question.

Interest in performance rewards is growing at an odd and unsettled time in American education. Standardized norm-referenced tests are in bad odor because they have defined academic performance in fragmented and often quite limited terms. Authentic assessment rests on a more complex vision of academic performance, but it is a much less well-developed technology, and thus would be less useful for making the sorts of decisions that performance reward schemes would require.

Finally, we have little evidence about how well the newer assessments represent students’ “understanding” of mathematics or reading, in part because this assessment technology is very young. It seems reasonable to think that performance assessments could do a better job than standardized norm-referenced tests, but researchers are only now turning their attention to what understanding might be, how it might be assessed, and what satisfactory assessment of understanding might be. But even if we had much more experience with such assessments, it is almost certain that testing experts still would disagree sharply about the nature of tests and their suitability for use in performance reward schemes. It is even more likely that disciplinary scholars would continue to disagree both about what is most important in their fields and about the nature of adequate performance in whatever they deemed most important, and that ordinary Americans would persist in differing about the sorts of academic performance that should be assessed.

Feasibility

Standardized tests would be the easiest assessment to use by far. They define knowledge in ways that are thought to be closer to established classroom practice than any of the newer and more ambitious assessments. The measurement technology is well developed and stable. These tests also are relatively easy for teachers and administrators to use, and they are familiar to students. They are not part of instruction and intrude on it only modestly. Educators also know how to administer them from much experience so that little new training would be required. And the tests are scored elsewhere; teachers’ administrative responsibility ends once the answer sheets are collected.

Authentic assessments would be much more difficult. One reason is that they define knowledge in more open and complex ways than their better established cousins, which greatly complicates test construction and evaluation of students’ performance. Another
reason is that the technology for these novel assessments is young, and much remains to be done, including learning about the assessments' validity and reliability. Until the technology was more mature, assessment would be an uncertain enterprise, and that would be a major problem in schemes that attached rewards and penalties to the performances being assessed. A third source of difficulty is that such assessments have to be scored by knowledgeable judges—that is, teachers—and that would be costly in time and demanding in expertise. Fourth, these assessments are intended to be integrated into instruction by incorporation both as instructional goals and as evidence of students' performance. But to use them well for either purpose, most teachers would have to unlearn and learn a great deal about academic subjects and about new approaches to instruction and assessment; knowledge of all three is in short supply.

Portfolio assessments seem a promising way to relate assessment to instruction, but they would be very difficult to integrate into many performance reward schemes. One reason is that portfolios comprise materials that are produced jointly and over time—by students and their teachers, parents, and peers. Hence, they may not be a valid measure of individual performance, and they may not reliably relate to other performances of the same individual. Another reason for caution is that because portfolios are built in an open process rather than composed on demand, they would be especially open to the influence of parents, friends, and teachers that one would want to exclude from any performance reward scheme. It would not be easy to use such assessments in ways that were both fair and believed to be fair, especially when serious consequences for teachers and students hung in the balance. Still another reason for caution is that unless all schools in a performance reward scheme had the same curriculum and the same criteria for selecting materials, portfolios could in some sense be incommensurate. And to the extent they were, ratings across schools—or even across classrooms within schools—could be an unsteady basis for making decisions about rewards and punishments. Efforts to create acceptable interrater reliabilities could be enormously time and energy consuming, and they still might not create adequate confidence in the results. Comparability among teachers and schools could be enhanced by carefully specifying the sorts of student work, the topics, and the like, but the more highly specified those terms of reference became, the more such assessments would resemble a districtwide take-home exam rather than a portfolio that allowed each student to present his or her unique strengths, perspectives, and modes of work.

Familiarity to parents, policy makers, politicians, reporters, and others concerned with schools is another dimension of feasibility. The more innovative an assessment was, the more difficulty such people would have in recognizing it as a valid and sensible instrument of education. The more intellectually ambitious assessments were, the more difficulty teachers and students would have, and the more students would do poorly—at least initially. As the recent debacle with the California Assessment Program revealed, innovative and ambitious assessments can be political dynamite, especially in a society in which cultural conflict over schooling is a favorite entertainment.

Cost

Standardized norm- or criterion-referenced tests are relatively cheap. Testing firms have been producing them for years, and can continue to do so for modest cost. Such tests also take little time to administer and are machine scored, thus holding labor costs down. But authentic and portfolio assessments are not being produced en masse; they are new and expensive to design—in part because in order to be usable, teachers need to be involved in design, field trials, and revisions. They also are much more expensive to use partly because they must be scored by well-educated observers and partly because they take a good deal more time than standardized tests.

At the moment, then, the assessments that many researchers and reformers consider least desirable are the easiest and cheapest to compose, administer, score, and use, whereas the assessments that many consider most desirable are the most difficult and costly to compose, administer, score, and use. School systems that contemplated performance rewards could be drawn to standardized tests for reasons of cost and feasibility while being drawn to newer assessments for reasons of desired instruction and results. Some standardized assessments, including the National Assessment of Educational Progress and some state assessments, have begun to add items that are intended to tap more higher-order knowledge and skill, but there has been little research on the extent or effects of these changes. There also is only modest evidence on how standardized test performance correlates with performance on more innovative assessments, but the extent evidence suggests a weak relationship.
That may suggest that the new assessments measure very different things than the older tests, but it may be only that the new assessments are much less reliable than the older tests.13

We know only a little about how more innovative assessments would work in performance reward schemes, for there are only a few recent efforts to use them—those in Kentucky, California, and Vermont. California's experience has taught Americans how politically difficult innovative assessments can be, but we have nearly everything to learn about the other difficulties that will come up and how school officials and policy makers can manage them. More experience in Kentucky and Vermont may improve both assessments and the capability to use them.

Though this discussion tells us little about the actual effects of any particular performance reward scheme, it reveals that the effects on students and teachers are quite likely to vary with the measures of performance. My discussion also suggests that effects should be considered not only in student achievement, but also in professionals' time and energy in assessment administration and scoring, the complexity of assessment, difficulty of implementation, possible political conflict, and cost. There probably would be tradeoffs among these. For instance, some school systems that adopted performance reward schemes might wish to encourage complex student performance, but would not because of costs or the demands on professionals' time. One set of issues for research would concern how school systems dealt with these problems and what effects their solutions had on the design, implementation, and effects of performance reward schemes.

An additional effect of performance reward schemes concerns state and local testing programs. American schools administer more tests than any other country known to researchers. Some are part of inherited local assessment programs, others had their origin in state "back-to-basics" efforts, and still others are tied to specific programs, as with Title I testing requirements. These assessments are numerous, often burdensome, and sometimes offer contrary guidance for instruction. Would a state or local performance reward scheme give more coherence to the schools' assessment program by focusing attention on the scheme's measures of performance and drawing attention away from other assessments, or would the other assessments distract from the performance reward scheme? Such interactions between performance rewards and other state and local testing would be an important focus for research on performance rewards.

Finally, even though performance rewards focus on school achievement, that is not the only thing that Americans want from school or that responsible teachers and administrators care about. Educators also value the effort that students make, irrespective of their test scores. The qualities associated with effort—hard work, determination, dependability, and the like—are attributes that many employers seem to value. Many parents and teachers also place a high value on discipline, by which they seem to mean some combination of respect for authority, obedience, and orderly behavior, and these too are qualities that many employers value. Finally, some teachers and employers recently report that they set a great store on such qualities as independence of mind and the capability to work with others. But if such social and affective features of schooling are essential to decent schools, they are more difficult to assess in a valid and reliable fashion than traditional versions of academic achievement. One reason is that many of these aspects of schooling are not obvious—discipline can be viewed and enacted in several different ways—and it would take careful instrumentation and analysis to fairly capture them. Another reason is that capturing them in any measurement effort would require much labor-intensive observation and many interviews.

Observing and assessing these social and affective features of schooling would greatly complicate the task of deciding on the quality of schooling because it would add more indices of quality and accomplishment and many puzzles about how to weight them. It also would add expense, time, and trouble to any scheme, so it is easy to understand why school systems and policy makers would prefer to ignore such matters. But one cannot teach or manage a school well without giving close attention to discipline, mutual respect, hard work, dependability, and the like. To the extent that a performance reward scheme reduced attention to these aspects and outcomes of schooling, then to that extent would such a scheme sooner or later contribute to a different sort of educational problem.14

Criteria of Success

However they defined and assessed the results of schooling, decision makers would have settled on only the type of school performance
they wished to reward. They still would have to decide on criteria of success: given any set of measures of performance, what results would be good enough to warrant rewards, or bad enough to deserve punishment? That turns out to be a remarkably difficult matter because any criteria of success must satisfy various other criteria. For example, any defensible criterion of success or failure should be educationally sound, which requires taking account of various complex conditions that bear on instruction. But it is no less important that criteria of success be usable, for performance rewards could not provoke school improvement unless the scheme was easy for teachers, students, and parents to understand. That requires clarity and economy, which could be at odds with the complexity associated with educational soundness. In addition, criteria of success must be demanding enough to raise performance, but not so demanding that they discourage students and teachers. Those criteria also must be unusual enough to be challenging but familiar enough to command broad assent from educators, students, and parents. Finally, criteria of success must be fair—an idea that is simple to write but knotty to decide.

Single or Multiple Criteria

The easiest way to decide what level of performance should be rewarded would be to settle on a single score on a single test. Schools with an average score of, say, sixty or above would be rewarded, whereas those with an average score of fifty-nine or below would not. Rewards would be decided on the basis of one school average at one point in time. The great advantage of such a criterion is its simplicity, but that also is its great weakness. For pass/fail is a rather crude distinction, and it would be difficult to decide on a single score that was educationally defensible. The original minimum-competency testing programs were simple: each state or local program set an absolute cut-off point on the test, and students with scores above it passed; whereas those below it failed.

Decision makers could instead fix several levels on a single test that distinguished higher and lower ranges of successful performance. Such an approach would enable school systems to make at least a few distinctions among acceptable work and perhaps sched-
Static and Dynamic Criteria

A deeper problem with any fixed criteria is that they treat school performance in static terms even though schooling is a dynamic process. Students achieve particular levels of performance at given points in time, but these are only snapshots at one moment in a much longer process of academic growth and social development. Though such snapshots are reasonable, students, parents, and concerned observers also focus on students' progress over time as on their rank order at a single moment. A scheme that pressed students and teachers to orient their work to set rank-order targets could have very different results than a scheme that was oriented to improved rates of learning. For instance, fixed targets could make it difficult to fairly accommodate the enormous variations in academic ability and achievement that are found in most schools or school systems. Officials might choose a target that was keyed to students' average ability or achievement in an effort to compromise out the differences, but average achievement could be too difficult for academically untalented students and too easy for very able students. If so, an average could be ignored by both sorts of students, or it could offer teachers incentives to ignore both sorts of students, or both. Decision makers might therefore try to set different targets for average, more able, and less able students, but that would be difficult to do in a way that was both fair and educationally appropriate. We have little basis in educational research or theory for deciding what such targets might be, and the explicit acceptance of different results for different groups of students could provoke a firestorm of protest.

Disadvantaged students present an especially significant special case of this problem. They do less well; on average, than students from more advantaged homes. Through what many observers would say was no fault of their own, these students have lower average test scores than their advantaged peers. If school systems used fixed and academically demanding criteria of success, it would be very difficult for schools that enrolled many disadvantaged students to ever do well enough to gain rewards, for the rank order of school average scores would closely parallel the rank order of school average parents' education. To base decisions about schools' success on such rank orders would penalize many poor students and their teachers for disadvantages that were not of their own making. These students and teachers could reasonably argue that they were playing on a field that was tilted against them, and grow discouraged about their prospects of ever doing well enough to be rewarded. That would defeat the purpose of performance rewards.

The only way that school systems could remedy this problem would be to adopt a dynamic approach to assessing school performance—that is, use a value-added approach to devising criteria for performance. To do so, school systems could adjust school or classroom average gain scores at each grade to remove the effects of non-school factors on achievement. The point would be to restrict calculation of gains to those that schools may be presumed to cause in any given period, rather than those that could be traced to non-school influences. Differences in how much schools added to student's achievement would be the key to rewards. The most efficient way to make such adjustments is to statistically adjust students' fall-to-spring gain scores for differences in their prior performance, thus restricting the basis of rewards to effects that schools exert during the academic year.

But such compensatory adjustments would not solve every problem of fairness. Many schools enrolling mostly disadvantaged students probably would make less progress, on average, than schools that enrolled mostly advantaged students, even after adjustments for initial scores. That is especially likely if assessments focused on the sorts of ambitious knowledge and skills that many American teachers are only weakly qualified to teach. What might school officials or policy makers do in this event?

One likely response would be to set more modest gain score targets for schools that enrolled disadvantaged students, effectively basing target gains on student body composition or initial scores. Or officials could assign schools to performance gain streams, and compare adjusted gains only within them. If schools could not compete across these streams, the appearance of differential gains and levels would be hidden. But streamed competition could create potent disincentives for professionals to attempt dramatic increases in student performance. If so, such adjustments could further damage disadvantaged students. For each school's performance was compared only to that of others in its own stream, schools could "win" more easily. There would be little incentive for teachers and students to try to outperform schools in higher streams, for the purpose of streamed competition would be to avoid such comparisons. Hence those in weaker schools might be kept in a sort of educational
children more effectively. But other analysts and professionals and parents in disadvantaged schools would be likely to argue instead that it was exactly the increased fairness, and possibly additional money, that enabled some schools to mobilize to weaken the association. To adjust the associations and perhaps reduce added monies could penalize schools for success and perhaps push them back into failure.

There is a troublesome paradox here. On the one hand, if the effective schools had changed the technology of instruction, why give them the handicap that they had when they were less effective? But on the other hand, changing the adjustments could make it more difficult for unsuccessful schools to change by reducing their incentive to do so—if they changed they would lose some advantages. A similar problem could arise if poverty got appreciably worse, for its effects might then also grow worse, leading some to argue for changes in the adjustments. The arguments about these issues could be methodologically complex and politically gruesome. But the issues that I have noted reveal that if educators did appreciably improve instruction and achievement, they would be rewarded with a really difficult problem: are performance rewards a transitional measure—a means to discover better approaches to education that then could be put into practice without additional incentives—or are they an essential continuing element of any improved instruction?

This discussion also reminds us of the tradeoffs between fairness and clarity. My account thus far suggests that the more fair a scheme was the more complex and costly it would become. But the more complex and costly a scheme became the less intelligible and useful it probably would become. The alternatives to streamlined measures of performance that I sketched earlier have some appeal as a matter of fairness, but they could add complexities to an already complicated matter. A performance-reward scheme of that sort could become the educational equivalent of Ptolemaic astronomy, with adjustments loaded onto adjustments until few could comprehend. Simple schemes are more likely to be unfair, but more complex schemes that are more likely to solve the problem of fairness also are more likely to be difficult to understand. Teachers, students, and parents could not respond constructively if they lacked a clear sense of how a scheme worked, what produced rewarded or punished scores, and what actions were likely to be productive.
The problem extends further. The schemes that I have discussed here would not work unless school systems collected sound evidence with which to adjust student scores for the effects of prior performance. In high poverty schools, such adjustments almost surely would require evidence on student mobility, for without knowledge of who was in school for the entire performance period, adjustments could not accurately be made. Schools with high mobility rates work under a greater handicap than those that have quite stable enrollments, partly because students who were only there for half the year would not have had the same opportunity to learn as those enrolled for the entire year. If such distinctions could not accurately be made, it would be difficult to have confidence in the results, yet few school systems collect such evidence. Most rely on very crude aggregate data collected for other purposes. Without sound evidence, no one could convincingly defend the fairness of decisions about success, rewards, or adjustments to scores, but collecting the evidence could be complex and costly, and encourage political conflict and legal challenge. The political and fiscal costs of defensible data bases could keep many systems from doing a respectable job, but if state or local authorities attached serious consequences for students and teachers to analyses of school effectiveness, sooner or later officials would be pressed—politically, legally, and perhaps by a sense of professional responsibility—to collect the evidence and do the analyses in ways that met standards of decent social science. Though it seems clear that dynamic criteria of performance are more sensible than static ones, there may be deep problems in defining such criteria and calculating the added value of schooling.

It also is worth noticing that such determinations could be politically difficult. Criteria of success and fairness are quite murky under the current regime, for schools embody many purposes, conceptions of success, and fairness. Such murky complexity has been unavoidable in an institution that has been assigned many different hopes and purposes, that has been the site of many compromises, and in which many accumulated layers of purpose coexist. Yet performance reward schemes could not work as intended without much greater precision about criteria of success and standards of fairness, hence they would place much more educational and political pressure on any criteria and standards. One large question for research on performance rewards is whether either the criteria or the schools could bear those pressures.

What Gains Would Be Enough?

Even if a school system settled on a solid value-added criterion of success, officials still would have to decide how much added value would merit a reward. Many probably would incline to relative criteria of effectiveness—for instance, stipulating that schools with adjusted gains in the top quartile would be rewarded, or that all schools whose adjusted gain scores were above the district average of gains would be rewarded. There is no shortage of such relative criteria, and in all such cases, the cut-off for success would be set in relation to some statistical feature of the existing distribution of gain scores. Being above average or in the top quartile has considerable face validity and seems to avoid the need to justify performance levels in substantive educational terms.

- From this perspective, decisions about how much of a gain warranted a reward would be a matter of selecting the proper reference group. Many advocates see performance rewards as a way to improve achievement for students at the lower end of the achievement distribution, especially students from disadvantaged circumstances. From that angle, the performance of advantaged or high-achieving students seems a reasonable reference for decisions about successful education. But there is no scientific or educational basis for determinations of this sort, and they could lock in mediocrity. If Podunk had low average achievement and little variance in annual school gains, then a school’s rise to the average of the local distribution of local gain scores would not be a terrific achievement. One might ask what there is about Podunk that makes its top quartile of gain scores a defensible criterion of effectiveness.

That returns us to the problem that relative standards seemed to solve by avoidance: what is the educational justification for criteria of success? Many Americans recently have been convinced that something is deeply wrong with public education, that standards are not only too low but also aimed at the wrong target. Reformers argue that schools have been teaching outmoded facts and skills rather than advanced knowledge and critical thinking. If we want only a modest improvement in Podunk’s low-achieving schools, perhaps the Podunk average out of quartile and gains is an acceptable standard. But if we believe that most schools in Podunk and America aim too low and at the wrong goals, then standards anchored merely to statistical features of extant score distributions leave much to be desired.
Relative criteria might be more appealing if they were tied to some better reference than the distribution of extant scores. If Americans could agree on substantive educational standards and if sound assessments that were tied to such standards could be devised, then students' progress toward those standards could be examined and compared. School average scores could be adjusted to eliminate the effects of initial achievement, and schools' contribution to students' progress toward the standards could be computed and compared. Such a scheme would combine value-added measures of improvement with educationally defensible standards of performance. This would leave the problem of deciding how much student progress toward standards would warrant the designation of successful, and that would re-present most of the problems that I have discussed earlier in this section, including differential progress. But sound external, generally agreed-on standards could offer substantive educational grounds for deciding how much progress was enough.21

Several difficulties stand in the way of this appealing formulation. One is that the very idea of such assessments is quite novel. American educators have just begun to devise the standards that would be required, and it would take much experience and investigation to refine them. It also would take extensive experience and investigation to devise assessments that embodied those standards and that could be defended as such—work that also has just begun. Much careful work also would be required to define criteria of success on such assessments that were both attainable and educationally defensible. Another difficulty is that efforts to devise more ambitious goals, standards, and assessments have stirred up terrific controversy. Some educators and many citizens reject Goals 2000, critical thinking, and related reforms. Much of Goals and its state counterparts may not survive the current educational warfare, and many states may not build a solid foundation for standards-based reform. At the moment, the criteria of success that combine the most appealing features in principle seem the most remote in practice.

Failing Schools

A last important issue concerning criteria of success is criteria of failure—that is, how should officials rate schools that did not win rewards? Should all of them be designated "unsatisfactory" or "failed"?

That seems unwise, for it would label schools as either outstanding or awful rather than also being satisfactory or indifferent. Additionally, if the criteria for success were set relatively high, a pass—fail approach could produce a politically unacceptable avalanche of failures, with the likely result that criteria for success would be abandoned or set lower. Perhaps only especially low-performing schools should be rated as unsatisfactory, with those in the middle left unclassified. Such decisions would be consequential, for designating schools as "failed" or "unsatisfactory" might do more damage than the label of success could do good. Decisions about where to draw the line for failure thus would raise all the issues concerning criteria of success that I just discussed, perhaps with even greater stakes. There has been little discussion of this matter.

Some commentators might argue that there should be no designations of failure or unsatisfactory performance, that systems should focus only on success. But that seems a dubious strategy, for there is a good deal of evidence that a large fraction of schools, especially in disadvantaged neighborhoods, would not be able to mobilize themselves to substantially improve instruction even in the best-designed scheme with the most attractive incentives.22 Devising and using a designation of unsatisfactory performance would be essential if state or local officials were to reliably identify the most difficult cases and try to improve them.

What might be sound evidence of failure? Performance rewards are designed to improve students' achievement, but the first thing to consider, especially in the early years of any scheme, would be whether schools' instructional capability was improving. One would expect changes in instruction and management to precede changes in achievement, even in improving schools.23 And some instructional changes, such as more ambitious curriculum and teaching, would be likely to initially depress achievement, but it would be foolish to penalize schools for trying to accept more demanding goals. Additionally, some schools might make substantial progress in instruction and management but not achievement, whereas others might make no progress on either front. In any reasonable situation, the first sort of school would be candidates for assistance in making more change, and the second would be candidates for a different kind of treatment. It would be unwise to use a single blanket designation like "failed" for schools that were so different and required such different treatment.
But if these suggestions make sense educationally, they could have troublesome administrative consequences. My proposal would require schools to collect evidence on change in management and instruction, along with achievement, and to monitor both, at least in dealing with the schools most in need of improvement. To be done well would entail careful, costly, and time-consuming investigations by capable researchers. That, in turn, would add more evidence, reporting, and analysis, further complicating the tasks of parents, teachers, administrators, and policy makers. Moreover, it would only be the beginning, for if experience and research are any guide, it often would take extraordinary work to rebuild or replace troubled schools.

Professional Capability

That point leads us to a central paradox of performance rewards. The schemes are designed to solve problems that arise from failures of professionalism, but they could not succeed unless professionalism improved. Advocates of performance rewards often slide over this point because they want to focus on incentives to use resources more effectively. They argue that the quality of teachers’ education, the quality of curricula and texts, the facilities in which teachers and students work, and the conditions of their work have improved greatly during the past century, all with no apparent corresponding increase in student performance. They point to many studies that report little or no relation between differences in schools’ educational resources and test scores, once students’ social status or beginning performance are taken into account. Advocates argue that unless there are incentives for professionals to use resources to improve students’ achievement, more resources will have no useful effect. Since few school systems offer such incentives, the chief task for school improvement is said to be rewarding or penalizing teachers for students’ performance. With such incentives in place, existing resources would be used more efficiently since teachers would either figure out how to produce the desired results or be replaced by others who could. Only when such incentives were in place could rational decisions about additional resources be made.

Even if one grants everything in that formulation, incentives would not be automatic. They could work only through people and their organizations, and if those lacked either the capability to respond constructively or the means to readily acquire it, incentives could have no useful effect. Hence anyone interested in how performance rewards might work also should be interested in the professional capacities required for a constructive response. At one level this would concern individual educators, for performance reward schemes would not work unless they were able to somehow learn how to do a much better job than they had been doing—barring the invention of some way for students to succeed in school without teachers. At another level it would concern school systems, for many schools that failed to perform successfully would need assistance to improve.

Individual Capacity

Speaking very crudely, there are two sorts of individual professional capacity—will and skill. Professionalism includes pedagogical skill, knowledge of subjects, and the like, but it is not just technical. Professionals also must want to help students learn; they must take responsibility for students’ work; they must care for students, respect their ideas, and believe in students’ ability to learn. Without technical capacity all the values and commitments in the world would be useless, but without those values and commitments all the professional knowledge and skill in the world would be impotent.

Many advocates of performance rewards tend to assume that professional capability is not an issue. They appear to believe that most professionals have the skill but not the will because the extant incentive structure in schools frustrates their efforts. There is one respect in which that analysis may be close to the truth, for there is appreciable evidence that time is poorly used in most schools and classrooms. The National Education Commission on Time and Learning recently concluded that “...even within the confines of a 180-day school year, reclaiming the academic day should, alone, nearly double the amount of instructional time in core curriculum areas.”

Many teachers spend a great deal of time preparing for instruction rather than teaching core curriculum subjects; many spend a great deal of time on holiday and other special events rather than teaching these core subjects; many take weeks to settle into the new year; many use the instructional time that they do spend quite
inefficiently; and a great deal of instruction consists of reviewing the previous year's work.25 There also is evidence that students who spend more time on academic learning learn more than those who spend less. If incentives for performance simply persuaded teachers to allocate more time to academic instruction, performance might improve for many students and schools.

But even that conclusion must be carefully hedged. One qualification is that most or all of any such improvement would be in what Americans call basic skills because most teachers do not presently have the knowledge or skill to teach in much more advanced ways. Another is that better use of time would not begin to exhaust the possible improvements that could be made in basic skill instruction because it would only more efficiently use teachers' existing skills. Researchers have recently delineated several features of effective direct or active instruction that go far beyond the efficient use of time. For example, teachers who work effectively have clear goals for instruction, they recurrently assess students' progress in lessons, they make strong and plain links between the goals that they set and the assignments, lesson design, and questions they ask students. They also maintain a crisp pace and a clear focus in classwork.27 These instructional actions rest on relatively complex knowledge and skills, and reflect significant professional norms. Research also suggests that only a modest fraction of teachers work in that way. For example, John Goodlad reported that most elementary teachers fail to make it clear to their students what is going on in the lessons so that many students are confused about what they are supposed to be doing, why they should do it, or both.28

A third qualification is that teachers would have much to learn simply in order to use time more efficiently at current skill levels. They would have to learn new methods of lesson preparation and format, pacing, classroom management, assessment of students' work, and much more. No small amount of teacher education would be required to help experienced teachers learn such things. Teachers also would have to learn to want to be more efficient—that is, to learn new professional values that would make their work lives much harder. Finally, they would have to learn that these changes would help students, which would require that they give much closer attention to students' work. Learning these things would be no mean feat, and there is no evidence that the capacity to offer the requisite pro-

fessional education exists today. Though we know roughly what efficient teachers need to do, few professional developers offer such education.

Even in a basic skills approach to performance rewards, then, there seem to be two levels of response. In one, changes in existing skills and knowledge might be sufficient to improve achievement for many students if teachers were motivated to use instructional time more efficiently and learned the required skill and knowledge. These things would not be easy, but they would not require an instructional revolution. Much greater capacity development would be required if teachers were to make further significant improvement in basic skills instruction, for the needed changes in instruction would be fairly radical.29

The professional capacity to respond constructively to performance records thus is not generic; the capacities required would vary with the performances that were to be assessed and the criteria of effectiveness that were set. In that connection I should note that in addition to the two alternatives sketched, there is a third, in which schools promoted intellectually much more demanding work. That would require much greater professional capability, for teachers would have to cultivate sophisticated knowledge of subject matter and pedagogy, and most lack such knowledge and skills. They would need extensive opportunities to learn if they were to respond constructively.30 There also is evidence that many teachers do not believe that most students can do intellectually challenging work. Hence they would have to acquire rather different professional beliefs about knowledge and students' capacities, and very different professional values to guide their teaching. These problems would be even greater for teachers who work with disadvantaged students.31

Advocates of performance rewards might object, saying that if strong incentives were attached to better student performance, teachers would somehow find ways to learn the necessary values, knowledge, and skills. On this view, greater demand would evoke an available but hitherto unused supply of professional knowledge and skill. But the evidence on professional development in education suggests that few suitable providers exist, even for improved basic skills instruction, and that the problem is most acute in schools that heavily enroll disadvantaged students. Hence, there is not only an appreciable lack of professional capacity to respond constructively
to performance rewards, but the lack is greatest in those schools in which improvement is most needed—and for those conceptions of instruction that are most ambitious.

Most public school systems that adopted performance reward schemes thus would not face the problem of mobilizing capacities already well developed but underused, but the much more difficult problem of developing and mobilizing barely existent capacities. That would be true even if one hoped only to squeeze more efficient time use out of the extant system, and it would be much more true for the more ambitious alternatives that I just discussed. State and local school systems have little experience with such work. Few seem to have any strategy for human resource development, except perhaps to identify areas like reading or mathematics for making professional development grants. Only a few have any capability for doing more. Most districts offer an array of brief and scattered professional development programs that reflect no particular strategy for human resource development and pursue no particular academic priorities. These considerations suggest that improving individual professionals’ capacity to respond to performance rewards would be a major task.

School Improvement

Improved professional capacity also includes changes in social organization, for the elements of capacity—values, commitment, and knowledge—are in some measure attributes of groups—that is, school and department staffs, and central office personnel. Milbrey McLaughlin and Joan Talbert argue that schools and departments within schools are professional communities that differ dramatically in their academic effectiveness in part because of differences in their professional norms. Anthony Bryk, Valerie Lee, and Peter Holland reach similar conclusions in their efforts to explain differences in secondary schools’ effectiveness. The capacity to respond to performance rewards should be understood as a problem of organization and culture as well as a matter of individual knowledge and values.

Organizational capacity is significant because spontaneously constructive responses to performance rewards could be expected only from a minority of schools that have extraordinary leadership or other unusual organizational resources. Studies of other school reforms suggest that many failing schools would be unable to correctly interpret the signals from a reform scheme, to fairly diagnose the problems and effectively devise means to improve instruction and raise students’ performance. Research on effective schools suggests that not many achieved and sustained high levels of student performance. Though a minority of schools would be able to spontaneously and dramatically improve their own instruction and management and students’ performance, it is very likely that a larger fraction either would make modest changes, inappropriate changes, or none at all. The existence of a group of schools that were mired in failure would tend to erode the credibility of any school system’s improvement efforts and leadership, and that would be a reason for the leaders of such a system to take steps to improve failing schools. Yet those officials also would have reason to be cautious about promising help, for few systems have the capacity to provide much help. Most state and local systems do not even identify failing schools (New York has just begun), let alone assist them in improving instruction. Few seem to have staff that are knowledgeable about school improvement. And if local or state school officials tried but failed to help weak schools improve, their claim to be professionally capable educational leaders would erode further.

Launching a performance reward scheme without a capacity to improve schools could be politically perilous. But devising such a capacity would not be easy, and failing at the effort also could be perilous. The causes of school success and failure are not well understood, and there is scant experience with or research on systematic school improvement. There is, of course, no shortage of schemes to improve schools, but there is relatively little reliable knowledge about what works, and even less about how to make it work in one place after it worked in another. Worse yet, the most credible research in this novel area suggests that effective and lasting school improvement is not only rare but quite expensive.

My account suggests that systematic learning from incentive schemes would be essential for even the hope of their success. School systems that instituted performance rewards should study the possible causes of school success and failure in order to learn from the results about how to improve individual schools. If school systems collected evidence on academic achievement only in order to decide which schools deserved rewards, they would have no solid basis either for commenting on why some schools succeeded and
and by students, teachers can succeed as professionals only when their students succeed as learners. Teachers therefore need their students to succeed. Of course, carpenters need wood and nails, but students have minds and volitions, and can influence teachers’ opportunities to succeed as professionals by the choices they make in working with teachers. Students can influence teachers in a way that wood cannot influence carpenters. Carpenters can produce results by themselves if they have the skills and knowledge of the trade, the will to work, and passable materials; for like workers in most occupations, carpenters address themselves either to inanimate materials or to ideas. But teachers, like therapists, organization developers, and other practitioners of human improvement, cannot produce results by themselves. Their art and craft are useless until students embrace the purposes of instruction as their own and seek them with their own art and craft. Of course, teachers need the will to work, and in this respect, they are just like carpenters and architects. But carpenters and architects do not require the will of their wood, paper, designs, and tools. Only teachers and their colleagues in other human improvement practices require clients who bend their own will and skill to the work.4

That fundamental difference among occupations bears on the mobilization of incentives for performance. It may make sense under certain conditions to focus external incentives exclusively on carpenters as a way to increase or improve the production of cabinets, for carpenters are the chief agent of cabinet production. But teachers are not the chief agent of student learning. Students are that agent, and teachers, however important, are ancillary. The instructional outcomes that advocates of performance rewards want to encourage are produced in and by students, with the help of teachers, parents, and others.

Teachers’ dependence on students is no theoretical matter, for high school teachers who are eager to plunge into Shakespeare regularly encounter students who want only accounting. Teachers cannot succeed—or even proceed—unless they reach some agreement with students about the ends and means of their work together, and since those agreements are regularly unmade, they often must be remade. Teachers are continually ensnared in negotiations about what students might do, and what teachers might do to secure students’ commitment. Many of these negotiations occur in the privacy of teachers’ minds as they try to anticipate students’ response to their...
next move or to possible lessons, but many others occur directly with students in class.

To teach at all is to engage in some such negotiation and calculation, but teachers are pulled in contrary ways. On the one hand, if their professional success depends on students' accomplishments, there seem to be powerful reasons to press students for ambitious work since more success for students is more success for practitioners. From this angle, there are potent internal incentives for teachers to encourage high achievement. Yet from another angle those vanish, for learning can be both risky and difficult, and more substantial learning is substantially more difficult. Teachers who demand high performance run greater risks of student failure, or resistance, or both. Since teachers succeed only through their students, they also have incentives to ensure some success for themselves by avoiding failure for students.\(^6\) One source of risk and failure is that many students seem uninterested in school learning. Another is that much learning entails unlearning: If one is to understand Newtonian physics, one must radically revise or abandon one's Aristotelian ideas about the motion of bodies even though the ideas are deeply rooted, serviceable, and difficult to cast aside. Another source of risk and failure is that learning often seems quite counterintuitive—for instance, that bodies in motion tend to remain in motion, or that we don't "see" things but rather interpret sense data in one of several ways that it could be interpreted. These and other difficulties create incentives for teachers to set low goals since a little improvement would be better than none.

I imagine teachers caught in a dilemma: should they aim low, risking mediocrity in return for at least some success for students and themselves, or aim high, risking failure in return for great improvement for their students and equally impressive accomplishments for themselves? I further envision the advocates of performance rewards proposing to intervene in teachers' dilemma by creating incentives for high achievement. If teachers and students worked in circumstances in which all other incentives for and against high performance were roughly balanced, then introducing a performance reward scheme would be predictably positive: the new incentives would tip the balance for many teachers. They would have reasons to work harder for improved achievement, and everyone would be a winner.

But in fact teachers and students work in schools, communities, and societies that have all sorts of other incentives and disincentives for student performance, and it may not be reasonable to assume that those are balanced in a way that would endow performance rewards with a decisive influence. For example, one very important source of incentives for performance is selectivity of students. Selective schools and programs capitalize on and create strong incentives for students to work hard and do well, and that eases teachers' dependence on students. Students accept that they should work hard and do well, so teachers need not become agents for mobilizing learners' commitment, nor do they need to lower expectations to keep students engaged. In such cases, the social organization of practice does the work of motivating students that teachers in many less selective schools must struggle with if they want students to work hard and do their best.\(^6\)

Selectivity generally has been weaker and less consistent in the United States than in most Asian and European systems, where students have been increasingly selected by ability and effort as the grades rise. By the middle or secondary grades, many academic teachers in nations like France, Singapore, and Japan are presented with increasingly able and committed students. There are, of course, some selective U.S. public schools, as well as top-track classes in many public schools; but even top-flight U.S. schools regularly present teachers with classes full of indifferent or hostile students, or with a mix of committed and uncommitted students. If those teachers want students to tackle demanding work, they must persistently try to persuade students to do so, taking major responsibility both for motivation and learning. It is much easier for teachers in such circumstances to reduce their dependence by adjusting their expectations to suit students' wishes. There is considerable evidence that many teachers, especially in secondary schools, adjust their expectations to suit students' wishes rather than pressing for high performance.\(^6\)

A performance reward scheme operating in such unselective schools and classes would offer teachers incentives to press students to do things that many students had other incentives not to do. Performance rewards would be working against the grain of the incentives arising from lack of selectivity. In itself that is not troublesome—teachers certainly should work hard to get students to do their best. But lacking any change in the larger structure of incentives for students, it is not clear that performance rewards would
have the desired effect. They might only heighten conflict within teachers and classrooms.

The issue cannot be decided in terms of selectivity alone, for several other incentive structures are at work. Teachers and students are influenced by the higher education institutions that admit high school graduates and the business firms that hire them. Those institutions' consumption patterns send signals concerning the qualities and accomplishments that they desire. American colleges and universities send mixed but generally weak signals about the importance of hard work and strong academic performance. One group of higher education institutions has very modest requirements: students need only a thin record of academic accomplishment in high school, often only a C average, to be admitted. Only high school graduation is required for admission in still another group of institutions, and not even high school graduation is needed in still a third group. These arrangements offer many students a second chance, but they also signal that high school students need not work hard in order to get into college or university. In those circumstances, it would be irrational for most students who aspire to higher education to work very hard in high school, just as it would be irrational for their teachers to press those students to try hard and do their best work. There are exceptions, of course; a small number of highly selective colleges and universities do have demanding admissions standards, and they do send clear signals to interested applicants about the importance of high performance in high school. But these schools are unusual.

A similar situation holds for the employment practices of most U.S. businesses. Few firms seem to ask for students' high school transcripts or references from teachers when considering them for employment. And even when firms do request transcripts, only a tiny fraction of schools supply them. The lack of employer interest deters students from thinking that grades, effort, or behavior count for jobs, and it deters teachers from thinking that their judgments about students could make a difference. If students can get jobs without even presenting evidence about their grades, school behavior, and teachers' evaluation of their work, it would be irrational for them to work hard, or for their teachers to press them for their best work. There are cases in which employers create relationships with specialized vocational high schools and hire many of their graduates, but these are exceptions.

These arrangements are unusual in the world. Universities in Japan, France, and many other nations lay great weight on students' performance in high school, high school leaving and university entrance exams, or both. If students wish to enter university, they must work hard in school and get good grades, make special preparation for the exams, or both. In many cases, the external examinations become the target of cooperative efforts by teachers and students to ensure that students do as well as they can. There are many troublesome features of such systems, including that they screen out able students who do not do well on exams and offer students no second chance at high attainment. But these systems leave little doubt in students' and teachers' minds that hard work and good school performance are important.

Employers in Japan, the former West Germany, and many other nations also pay close attention to students' secondary school records in hiring decisions. They routinely review transcripts and teacher references when high school graduates or early school leavers apply for jobs. In some cases, schools and employers work closely to place students in apprenticeship or regular work situations. Teachers know these things, as do students. It is understood that students who do not apply themselves and behave decently in school will have difficulty finding good apprenticeships or jobs. There are important rewards for academic effort and good behavior, even for students who have no ambitions for further education.

Given the broader incentive systems in which U.S. schools operate, schools could not adopt a performance reward scheme with much confidence about its effects. Teachers might be motivated to try to persuade students to work hard and do well in spite of the contrary signals from higher education and work. But lacking any change in the structure of incentives for students from firms and higher education, performance rewards might only heighten teachers' conflict, pushing them to push students in the opposite direction than students were being pushed by incentives from firms, higher education, and peers who knew the score on such matters.

Once again, the issue cannot be decided based only on the signals that firms and universities send students and teachers, for still other incentive structures bear on schools. Students and teachers also are influenced by broader attitudes about education, which are broadcast by the adults they know, as well as by newspapers, television, and other agencies. If these agencies sent generally strong
signals that hard work and high achievement would benefit students of all sorts, perhaps the lack of clear signals from higher education and firms would be somewhat offset; if so, performance rewards might tip the balance toward effort and achievement inside schools. But many send quite contrary signals. Americans are deeply ambivalent about intellectual work. Attitudes and values vary, of course—The New York Times and Public Television are different from USA Today and the Fox Network—but we have long been inclined to value experience over formal education, and to value practical rather than intellectual content within formal education. Teachers are not held in high esteem, and a large fraction of those who enter the profession still do so with the idea that it will be a backup in case a better job or marriage does not materialize. There is little on commercial television or radio that supports academic learning, and much to suggest that it is irrelevant or useless. Adults further report that they value learning to get along, and job-related knowledge and skills more highly than academic learning. For instance, 81 percent of the respondents in a recent Gallup poll said that the “chief reason” people want their children to get a formal education were job opportunities, preparation for a better life, better-paying jobs, and financial security. Only 15 percent said that the chief reason to get a formal education was to become more knowledgeable or to learn to think and understand. Americans also seem to act on these beliefs: relatively few mothers report working closely with their children on academic tasks or offering support for hard work and success in school.

Intellectual work and academic accomplishment are more highly regarded in many other societies. In Japan and China, for instance, parents take education very seriously and hold teachers in high esteem. Investigators report that Japanese and Chinese mothers also play a central role in their children’s academic work, encouraging children, working closely with them on assignments, and creating an environment conducive to learning. Japanese and Chinese mothers also seem to hold higher standards for their children and to have more realistic evaluations of their achievement than American mothers. Childrearing and adult values in these countries seem more conducive to successful schooling than in the United States.

In fact, America is distinctive for its deep divisions over schooling. We debate the value of intellectual independence versus obe-
dience. Many Americans want schools to teach only basic skills. More and more campaign for home schooling to save children from school’s “propaganda,” or they stage political shoot-outs over whether students should read Catcher in the Rye or The Adventures of Huckleberry Finn, or they argue about whether evolution, phonics, or critical thinking should be taught. Such matters barely surface in other nations. Our distinctive attachment to personal autonomy contrasts with more cooperative and deferential behavior in other societies, where people seem preoccupied with how they can fit in, work with others, and advance common values.

No one knows whether performance reward schemes would tip the balance of incentives toward high performance or would pit teachers against students by pitting school incentives for stiff academic performance against a broad array of social, economic, and cultural incentives for weak performance. But the abundant indirect evidence suggests that performance rewards would not have a generally positive effect because so many other incentives encourage low performance. That is not a reason to avoid incentives for performance, but it is a reason to design such schemes in light of the situation. Alternative designs might include schemes that mobilize incentives for students and teachers at once, rather than teachers only. One example would be direct incentives to students and teachers for improved student performance. Another would be instructional designs that focused teachers’ and students’ attention on the quality of students’ work in relation to demanding academic standards. Still another would be competitions among groups of students for rewards that they would value. In any such scheme, it would be critical to design things so that the rewards were congruent, so that teachers would have reason to be resources for students’ efforts to achieve. Another alternative design might include mobilizing broader social, educational, and economic incentives for student performance. One example would be persuading businesses to pay attention to students’ grades and teacher references, and to let schools know that they were doing so. Another would be to persuade higher education agencies to adopt higher standards for admission. These alternatives recognize that the motivation for school performance has at least as many sources outside schools as inside them, and that one element of good policy design would be to get the several influences working in the same direction.
Rewards and Penalties

In a perfect world, one would not need to discuss rewards and penalties. Teachers would do the right thing because they could do nothing else. But rewards and penalties are essential in an imperfect world, for they help us decide what is most important out of all that we need and want, and they help us influence others’ decisions. Rewards and penalties also can encourage us to cultivate or mobilize the capacities required to perform those tasks. The question is not whether various rewards and penalties are useful—they are essential—but what sort are best suited to improving student achievement.

Six assumptions, roughly derived from the discussion thus far, guide my response. The first is that we know only a little about the central issue: what suitable and serious rewards and penalties might be, and how to distinguish them from their too-trivial or heavy-duty brethren? That distinction would be critical in designing any scheme, but experience with and research on incentives in education are both quite modest. Much would have to be learned before we would have anything like sound design principles in this matter. It seems more appropriate to offer ideas that might guide the design of a scheme than to specify a design.

A second assumption is that one cannot intelligently discuss rewards and penalties apart from the other chief design elements of a scheme, for rewards and penalties would not operate in isolation. The design elements would specify the situation in which rewards and penalties would operate, and thus would offer at least a rudimentary basis for considering how they might work. My remaining assumptions fill in the most important design elements. One is that although incentives for students are as important as those for teachers, they do not substitute for each other, for teachers’ perceptions and interests are far from identical with those of students in most American public schools. Incentives for teachers to improve student performance should be complemented by incentives for students to improve their own performance. Though those incentives might be different, they should be aimed at the same results and designed to encourage the same sorts of academic behavior.

Another assumption is that any serious performance reward scheme would aim at higher level knowledge and intellectual skills, which could include demanding versions of traditional instruction.

Still another is that such a scheme would sharply increase uncertainty, and that would become more troublesome as learning goals became intellectually more ambitious. Such learning goals are more complex, open more room for disagreement than those associated with basic skills, and thus create more uncertainty in and around classrooms. Advocates argue that incentives for results would have a salutary effect on learning because they would focus attention on results and thus produce better performance. But that assumes a stable and well-known instructional technology, which does not exist for intellectually demanding work. Though many teachers care deeply about results, they often focus on instructional processes partly because they find learning to be so uncertain. It is, for example, often difficult to know when students know something: those who seemed to know two-digit multiplication on Monday often seem not to know it on Friday, those who seemed to know it when problems were written one way seem not to know it when they are presented in another. It is much easier to tell students to work two-digit multiplication problems on a worksheet than to probe their knowledge deeply—that is, ask them to solve the “same” problems in several different forms at several different times. A scheme that offered potent incentives to focus on results would be likely to increase exactly those uncertainties that most trouble most teachers. Lacking stable, effective, and well-understood systems of instruction, many teachers would respond to performance rewards by fixing on process indicators that seemed to be proxies for the required results, or mechanically imitating instructional approaches that they thought would produce those results, or both.

A last assumption, closely related to these comments, is that any performance reward scheme that aimed higher than low-level skills would require fairly extensive teacher learning. Rewards and punishments should not only encourage teachers and students to boost student performance, but should also encourage teachers to learn what they would need to know in order to do so.

There are at least several ways to design schemes that would fit with these ideas, and we are far from knowing enough to specify the One Best Design. But one way to make close connections between the performances for which students and teachers would be rewarded would be to make students’ academic work the focus of teachers’ work—that is, to organize schooling so that teachers devoted considerable attention to understanding students’ work and
learning how to improve it. Linda Darling-Hammond terms this "learner-centered instruction." Such an approach could create a basis for linking teachers' and students' learning by making students' work the agenda for teachers' learning. The teachers' curriculum would begin with students' work, with teachers' efforts to improve it, with the subject matter in which students' work was situated, and other related matters.

Teachers' work on students' work thus would focus partly on efforts to improve students' performance, dealing with such questions as "how good is this answer?" and "why is it as good as, or better than, that answer?" and "how could I have helped this student do better—and how do others do it?" One result of such work would be examples of desired performance that teachers could use in their work together, as well as in their work with students. In order for such a scheme to work, teachers would have to agree on results and judgments of quality, and come to some common understanding of acceptable variation. Teachers could not do much of that without opportunities to work together on students' work, on their understanding of it, their views of its quality, and their grasp of the underlying material.

Those ideas also imply the existence of common assessments that teachers and students also would use to guide their work. And that in turn suggests that there would have to be something like a common curriculum—at least agreement on the content to be covered, as in the Advanced Placement program. For if teachers did not have a common range of topics to work on, it would be extremely difficult to frame common assignments, to develop common ideas about quality, and to develop common understanding of how to improve teaching and learning. Without such things, no scheme could have any broad validity or application.

The point in such an approach would not be to "restructure" schools, but to revise instruction so as to create extensive opportunities for teachers to focus on students' work and its improvement, and, to that end, to improve teachers' knowledge of subject matter, teaching, and learning. Situating those collective efforts to improve students' performance in instruction would create a basis for rewarding teachers and students jointly for improved performance, but it would initially increase uncertainty and teachers' workload. If done well, it could help teachers learn what they would need to know in order to deal with the increased uncertainty of a performance reward scheme, by building a foundation for common judgments about the quality of students' work and what knowledge counts most. Such an endeavor also could reduce mechanistic instruction and create extensive support for teachers' learning.

A scheme of this sort could mobilize many internal incentives for teachers and students. For instruction that enabled students to learn more, and more effectively, would generate powerful satisfactions for teachers and students, and thereby incentives to sustain the effort. Such internal incentives would have to be central to any serious performance reward scheme, for teachers and other practitioners of human improvement depend on clients for their success. Lacking a way to mobilize those mutual satisfactions of teaching and learning, it is difficult to imagine external rewards and penalties that would be potent enough to produce high levels of performance.

But even if such internal rewards and penalties became central to teachers' and students' work together, they could not carry the entire weight of the reforms discussed here. For the transition to a more intellectually ambitious and performance-oriented sort of instruction would be difficult, unsettling, and in some important ways quite unsatisfying for teachers, not to mention students and parents. More palpable and less subtle rewards and penalties also would be important. Salaries, working conditions, and how superiors respond to their work all matter to teachers, and would be suitable material for performance rewards. It probably would make sense to combine various sorts of rewards—financial and professional, individual and collective—rather than focusing entirely on one or the other. Suitable financial rewards could include: free tuition, money for instructional improvement, and salary bonuses, whereas suitable external professional rewards could include career advancement, public recognition, improved working conditions, and the like. No medium is pure—money rewards for good work with students would be more than money, for they also would signal some professional recognition and might lead to opportunities for advancement or improved working conditions. But the emphasis among these can differ considerably: One reason for a mix of rewards is that it could reflect the rewards of teaching. Another is that professional taste varies—many would prefer one sort of reward to another, but few would flatly reject either. Still another reason to mix rewards is that the scheme I have sketched would require much collective work by teachers, and it could be damaging to assign
individual rewards for collective effort. It probably makes sense for the balance among these rewards to be locally designed, or at least to be open to local redesign.

Finally, one must decide whether the absence of rewards is a penalty and, if so, whether it is a sufficient penalty, especially in the schools most in need of improvement. If not, what penalties would be suitable? The answer to the first query depends partly on the perceived fairness and legitimacy of the scheme. If both were high, and if knowledge of the scheme were broadly and thoroughly available, then the absence of success and rewards probably could seem to be a penalty. But such penalties seem likely to be sufficient only for schools that could relatively easily be improved, and those probably are a minority of the schools that most need improvement. Hence, additional penalties should be considered, including:

- Pressuring principals and teachers to improve
- Replacing the principal and other school leaders
- Teachers losing assignments or tenure and replacements being hired, handing over the school to an external improvement team, or closing the school and reopening it as a new unit with new leadership, staff, and mission

These alternatives remind us that the schools most in need of improvement would be least affected by the absence of rewards or the presence of penalties because they would have the least capability to read the situation, to consider approaches to improvement, and to take effective steps. That, after all, is one of the chief reasons that such schools are poor. Hence, we confront the likely inefficacy of incentives alone to provoke action that would broadly improve schools. My analysis suggests that externally provoked and operated school improvement also would be required.

Conclusion

I have repeatedly alluded to the limits of this little exercise in design. A scheme that moved schools toward a focus on results would increase teachers’ uncertainties, especially if it also moved toward intellectually more ambitious conceptions of results. Many classroom problems for teachers and students would predictably follow from these changes, and they would be likely to generate pressures for simpler and more familiar assignments and assessments. That was the fate of the 1950’s curriculum reforms, and as I write, reform in California, South Carolina, and other states is under attack on such grounds among others. A movement to press schooling back toward basics has been gathering force for a few years, but the further school systems moved in the direction of such simplification, the more performance reward schemes would be trivialized or regressive.

These matters could profoundly affect the design of any performance reward scheme, but they are not matters that designers could control. Designers might at best devise measures of performance and rewards and penalties that could ease the probable political pressures and perhaps restrain them from corrupting the design. One way to do that would be to devise assessments that covered several conceptions of knowledge and learning, and another would be to design rewards and penalties in consultation with parents and others interested. Performances and rewards should be understood as much in political as educational terms, for in America no scheme to improve performance will operate well educationally unless it works well politically.

Discussing the politics of performance rewards calls to mind their economics. Many advocates find these schemes appealing because they promise cleaner, more efficient, and perhaps leaner school systems. But I would mark that down as one of several central issues for research. My analysis suggests that serious performance reward schemes would be unlikely to work without a considerable additional investment in professionals’ learning and school improvement, and those would entail more administrative responsibility and costs. Performance-oriented school systems might be differently organized and better, but they seem unlikely to be less expensive, or to have less potent central offices. The political and fiscal price of greater effectiveness may be steeper than many imagine.

Though this essay has covered a considerable waterfront, I have so far said nothing about how much improvement would be enough. I suggested criteria of success that were linked to substantial educational goals and standards, but no one knows how much scores should improve. Should state or local schools expect modest gains? Large gains? Should the entire distribution of achievement rise, should we aim for less inequality overall, or both? Research sheds little light, either on what reasonable expectations might be or on how much time improvement should take. These would be critical issues
in the public as well as the professional evaluation of performance reward schemes, but researchers and professionals have barely made a start on them.

That point, and many others like it show that performance rewards in schools are mostly unknown territory. A polite way to put it is to note how much research could improve understanding of performance rewards. We don't know enough about incentives or school improvement to estimate, for example, the size or types of incentives that would be most compelling for teachers. In a rational world, such weak knowledge would incline public officials and others interested in creating opportunities to systematically learn from experience. It would, for instance, be appealing to devise some sort of quasi-experimental or planned variation design in order to improve learning from innovation. But such things are costly, politically troublesome, and methodologically difficult in the best of times. And our own times is distressingly irrational, marked by expanding ideological conflict and contracting budgets; it makes experiments and other efforts to systematically learn from experience seem a bit remote.

Notes

1. This paper was written in connection with a forum on Incentives and Systemic Reform, sponsored by The Consortium For Policy Research in Education (CPRE) and The Pew Forum on Education Reform. I am grateful to Marshall Smith and Richard Murnane, whose dissatisfaction with an earlier version prodded me to rework it. Murnane also helped me think about how to revise, as did Tony Bryk, Susan Fuhrman, Helen Ladd, Jennifer O'Day, and Janet A. Weiss. My thinking about problems of professionalism was improved by comments from Brian Rowan and Joan Talbert, and Gail Baxter had helpful comments concerning assessment. My understanding of several issues was improved by discussions on performance-based approaches to school reform that Helen Ladd organized at the Brookings Institution. Of course none of these people are responsible for my views.

2. The literature on trends in student performance is very mixed, and researchers disagree. Eric Hanushek (Hanushek et al., 1994, chapter 3) argues that during the last three decades student performance on standardized tests has declined or, at best, remained unimproved. In their study of historical trends in literacy, Lawrence Stedman and Carl Kaestle qualify critics' and reformers' claims of declining test scores by noting their failures to account for changes in test-taking populations and contradictory evidence (Stedman and Kaestle, 1987, pp. 18-23). They are inclined to think the scores have held roughly steady despite large increases in the enrollment of educationally disadvantaged students. Gerald Bracey takes much more sharp issue with claims of educational failure that cite declining test scores (Bracey, 1992). Like Stedman and Kaestle, he believes that demographic changes need to be factored into longitudinal comparisons of average student test scores. But Daniel Koretz argues that achievement has declined since the 1950s, even after compositional changes in the school population have been taken into account (Koretz, 1986; Koretz, 1987).

3. Linda Darling-Hammond argues against many sorts of accountability on the grounds that rewards and sanctions tied to student performance exacerbate existing educational inequalities and preclude meaningful inquiry into teaching and learning at school and classroom levels though she supports certain other sorts of accountability (Darling-Hammond, 1994a, pp. 14-16, 23-25). But Eric Hanushek asserts that incentives linked to student performance are "the best hope for getting on a path of long-run improvement" (Hanushek et al., 1994, p. 88). Chapter 6 of his book offers a taxonomy of performance incentive schemes, including schemes to reward or penalize teachers based on student performance.

4. This sort of scientifically augmented state administration could be compatible with market solutions. States could operate schools and permit new schools to enter the market and permit consumers to choose another school if theirs was low performing. But most accountability schemes do not go that far—they rely on state administration augmented with evidence on effectiveness and some sort of penalties and rewards.


6. Passing the National Board For Professional Teaching Standards exams is one exception to this generalization.


8. Gail Baxter, Anastasia Elder, and Robert Glaser provide evidence of a strong positive relationship between student scores and sophistication of thinking activity on a science performance assessment (Baxter et al,
unpublished manuscript, pp. 14–26). Norman Frederiksen reviews a number of instructional and assessment technologies that support higher-order thinking by stimulating real problem situations (Frederiksen, 1994, pp. 550–558). Lauren Resnick acknowledges the praise that individual performance assessments receive for promoting higher-order knowledge and skills, such as those identified by the authors just cited, but she notes the absence of evidence for performance assessments’ collective construct validity (Resnick, 1994). She attributes the absence of such evidence in part to a lack of consensus over the content to be measured by performance assessments (pp. 522–525).

9. Lauren Resnick provides a hypothetical sketch of a portfolio and performance assessment system that illustrates the greater capacity of portfolio assessment for representing a variety of student work (Resnick, 1994, pp. 513–516).

10. Test items are selected by first sampling topics, problems, and other materials from a variety of texts and curricula, developing pools of test items, trying them out, and selecting only items that produce normal distributions of responses. That rules out items that many students get right or wrong.


12. Ibid.

13. Gail Baxter, Richard Shavelson, Susan Goldman, and Jerry Pine find only “moderate” correlations between student performances on standardized multiple choice science tests and more innovative hands-on science investigations (Baxter et al., 1992, p. 12; Shavelson et al., 1992, p. 25). In addition, they find stronger correlations between standardized tests and measures of general cognitive ability than they do between innovative assessments and ability measures. Taking these two findings in combination, they conclude that standardized and innovative assessments measure different aspects of science achievement, with the former being more related to students’ general knowledge and skills (Baxter et al., 1992, pp. 11–13; Shavelson et al., 1992, p. 25).


16. Dallas has such a scheme, and it is terrifically controversial. For one account, see Helen Ladd and Charles Clotfelter’s chapter in Ladd, op. cit: Linda Darling-Hammond and Ernesto Cortes, in personal communications, argue that the Dallas scheme is fatally flawed, among other reasons because the most disadvantaged schools are effectively excluded from the scheme.

17. One bit of additional support for value-added approaches to criteria of success is evidence that students in high-poverty schools seem to progress at roughly the same rate during the school year as students in more advantaged schools—the rates diverge during the summers (Heyns, B., 1978). This study suggests that using fixed criteria would bestow advantages on schools in privileged communities that would not accrue if districts used gain scores that were adjusted for inherited status, prior scores, or both, and did not include summer learning. But Heyns’s data are nearly twenty years old, and more studies of relative gains would be needed to prove the issue further.

18. Whether that turned out to be the case would be an important issue for research on performance reward schemes. But many commentators expect that one effect of setting more demanding educational goals and standards—à la Goals 2000—is that more advantaged students would be able to do much better on them than disadvantaged students, both because of the educational consequences of social and economic disadvantage and because instruction in disadvantaged schools often is less effective than in more advantaged schools.

19. Whether schools were able to become much more effective in a stable fashion year after year would be one of the most critical issues for research on performance rewards.

20. Elmore, Fuhrman, and Abelman, op. cit, report that Mississippi recently adopted such a scheme and that it is quite difficult for teachers and parents to understand. Few seem to comprehend why schools that gained rewards did so, or why others did not. Mystification and mistrust seem to mark responses to that state’s scheme, which do not bode well for its effect on practice.

21. That appears to be roughly the approach envisioned in Title III of the Clinton administration’s central school reform legislation, Goals 2000. That statute invites states to adopt “internationally competitive” instructional goals and content and performance standards. Once those were developed and in place, assessments keyed to them would be devised or selected. Title III of the statute seems to urge educators to use assessment of (at least) students in disadvantaged schools to decide about movement toward new state standards. But, as I write, Title III is dead.

22. There is some evidence from both the South Carolina and Dallas plans, and from the Chicago reform, that quite a few schools would fall into and remain in such a category. See Ladd and Clotfelter, op. cit.


24. Hanushek et al., op. cit.

29. Even such work would not be easy despite the focus on basic skills. Though most observers agree that classroom work is generally oriented to basic skills instruction, and that most teachers have not been educated to do much more, few have been educated to do such work very effectively. Doing so is considerably more demanding than ordinary classroom work. As a result, efforts to create more effective basic skills instruction run into appreciable obstacles. For example, Jane Stallings, reported in studies of the implementation of a direct instruction program in northern California, that students’ achievement improved as long as teachers were being trained but that scores slipped once training was withdrawn.
31. For example, teachers of disadvantaged students hold their students’ academic abilities in lower esteem than teachers of advantaged students (see U.S. Department of Education, 1993, pp. 84, 87), and many teachers attribute their students’ weakness to family circumstances. The frequently implied conclusion is that schools and teachers can do little to help. Increasing professionals’ capability to respond constructively to performance reward schemes would require change in such beliefs, for teachers could hardly boost students’ achievement if they thought them incapable of learning very much, or if they assumed that teachers were not responsible for producing better results, or both. Skill and will work together when they work. To be instructionally effective, teachers’ sense of responsibility for students would have to be linked to an extensive repertoire of instructional actions and to criteria for judging the results. For instance, in order for educational professionals to take responsibility for students’ work, they would require high internal standards for the sorts of things they should do to interest students in the work, to maintain students’ commitment, and to encourage students to work hard. If such teachers had students who failed to measure up, they would revise their approach and try again. If they failed to notice a student’s problem that later came to light, or if they noticed it and failed to take appropriate action, they would question their judgment or motivation, and try to figure out how to correct the problem.
33. See, for example, Bryk, A. and Lee, V. (1993).
35. Bryk and Lee, op. cit, chapter 11.
36. Bryk et al., op. cit.
40. The need for systematic learning also follows logically from the arguments made by advocates of performance rewards that we lack solid knowledge concerning the links between school resources and school achievement because under the present regime there are insufficient incentives for performance. Only when such incentives were in place would there be fertile ground in educational reality in which valid knowledge about instructional improvement could thrive—whether in the form of craft knowledge or a well-specified production function. These ideas suggest that the existence of performance rewards would be an occasion to create the knowledge and skills that would be needed to operate performance reward systems and help failing schools improve.
41. Some readers may say that nearly all occupations try to improve human and that nearly all have managers who work on other people and try to produce improvements. But consider the nature of the improvements and the means by which they are cultivated. Surgeons do not try to make their patients into apprentice surgeons, nor do salespeople try to improve their clients’ capacity to sell vacuum cleaners or encyclopedias. Moreover, most managers assiduously avoid trying to help their subordinates become managers, let alone better managers. In occupations like sales, physical medicine, and many branches of management, practitioners typically strive for a distinctive sort of result: items sold or manufactured, profits earned, bones and organs repaired, and the like. Betterment of clients’ minds and souls, or improvement of their knowledge and skill are subsidiary at best, and typically either irrelevant or merely decorative. In contrast, teachers succeed only by helping students acquire some elements of their own specialized expertise: knowledge of a subject, skill in communicating about it, a repertoire of strategies for solving problems, and the like. When psychotherapists succeed, it is typically by helping their patients acquire elements of their own distinctive therapeutic expertise: insight into emotional problems, understanding their sources, skill in noticing symptoms, and a grasp of the barriers to improvement. Only teachers and practitioners in sister trades must cultivate their clients’ capacities to become adept practitioners of their own improvement, for only in these professions must clients become such practitioners in order for professionals to succeed. Hence, though there are elements
of human improvement in many modern occupations, there also are important differences between human improvement professions and other occupations in which people are processed.


46. Bishop, op. cit.

47. James Rosenbaum and Takehiko Kariya conclude that institutional linkages between Japanese employers and high schools increase the influence of academic achievement (grades) on work-bound graduates’ chances of attaining desirable jobs; consequently, they create clear incentives for students to achieve academically (Rosenbaum and Kariya, 1989, pp. 342–358). Rosenbaum and Kariya also find that Japanese junior high students develop their educational and occupational plans in direct response to their academic achievements (Kariya and Rosenbaum, 1987). The centrality of academic achievement to their planning results from the education system’s exclusive reliance on achievement for high school admission selections.


50. Lynd, R., Lynd, H., op. cit; Cusick, op. cit; Powell, Farrar, and Cohen, op. cit.


54. Linda Darling-Hammond’s proposals for learner-centered instruction (see her essay in this volume) are a promising case in point, for they recognize that teachers’ success is intimately linked to students’ success—that is, she assumes that most teachers have professional incentives to help students learn if they have the wherewithal to do so. It is not clear whether her proposals take account of the risk and difficulty that more ambitious instruction holds, and of the ways those things can push teachers to accept modest success through modest standards for students.


56. Linda Darling-Hammond, op. cit.

57. See, most recently, Hanushek, E., op. cit.

References


Cohen, D. K., and Spillane, J. P. “Policy and Practice: The Relations


