Incentives and Equity under Standards-Based Reform

JULIAN R. BETTS and ROBERT M. COSTRELL

Standards-based reform is a strategy that includes specifying what is to be learned, devising tests to measure learning, and establishing consequences of performance for students and schools (for example, setting cut scores for grade promotion and high school graduation). The goal of this strategy is to raise student performance across the spectrum, especially for students from those schools, often heavily minority, where expectations are chronically low. The point is to alter incentives and change the behavior of students, teachers, administrators, and parents in a way that improves learning.

Popular support remains strong for standards-based reform, according to national polling data as well as local data in the states implementing this strategy. For example, a recent poll in Massachusetts, which is implementing one of the more rigorous sets of exams (effective for the class of 2003), indicates that 70 percent of the general population favors graduation exams. Support is slightly more emphatic from urban than suburban respondents, and somewhat broader (75 percent) from those with income under $25,000. When respondents are asked if they would still support the exams should 25 percent of students in their communities fail on the first try, support remains unchanged overall at 70 percent and rises to 81 percent among those with income under $25,000.

 Nonetheless, vocal, if not yet necessarily wide, opposition has emerged in several states, in the run-up to full implementation of standards-based reforms. Objections fall into different categories. One source of discord concerns the content of what should be learned. The battles of the mid-1990s over national
content standards in history and English, and more recently in science and mathematics, have had their counterparts in the states.\textsuperscript{3} Despite continuing conflicts, certain broad (if not universal) agreement can be obtained in basic content areas (at least mathematics and English). The focus here will not be on content disagreements, but on disputes over testing and cutoff scores. However, at least some of the more vocal opposition to testing is based (if not always explicitly so) on unresolved disagreements over content standards, because tests give force to the content standards.

Opposition to testing-with-consequences is based on a simple, fundamental fact of life: Almost any change creates winners and losers. For example, technological progress has always had its losers, from the hand-loom weavers to the buggy-makers to current-day bricks-and-mortar retailers, computer illiterates, and those of low cognitive skills. The technologically caused losses of those with low cognitive skills over the last two decades have driven much of the standards-based reform movement. So, too, may standards-based reform create losers (at least in the short run) in the attempt to create more winners from technological progress. The fact that there are losers, along with winners, is not, in itself, a compelling reason to roll back the standards any more than it would be a reason to try to halt technical progress (by, say, shutting down the U.S. Patent Office). Instead, the nature of the losses must be examined and an appropriate set of policies must be crafted to minimize them.

The most obvious potential losers are those who may not meet the standard, and who may not earn a high school diploma as a result. But this is only the beginning of the analysis. For example, whether the failure rate rises as a result of sorting or whether it also reflects adverse incentive effects makes a great deal of difference. The distinction is important both for evaluating the costs of increased standards and for focusing policies to mitigate costs. Similarly, distinguishing sorting and incentive effects among the winners from various points on the educational spectrum is important.

Standards generate a mix of sorting and incentive effects. How are incentives altered by standards-based reform, for better or for worse, to encourage or discourage achievement? What are the trade-offs between some students’ losses and others’ gains, in learning or income? Do these trade-offs adversely affect equity, as opponents to standards-based reform often claim? Or is equity enhanced by raising standards in schools attended by disadvantaged students? Why does some of the most vocal opposition often come from the most advantaged districts? Finally, and most important, what steps can and should be taken to minimize the losses and spread the gains most broadly from standards-based reform?
The Economic Theory of Educational Standards

The economic theory of educational standards attempts to elucidate the likely effects on learning incentives and economic outcomes by means of a simplified model. Economic theory is applied to the subject of standards because economics offers a well-developed framework for the study of incentives, which lie at the heart of standards-based reform. It also offers a systematic method for identifying likely winners and losers, and, more important, the reasons behind and nature of the gains and losses. Finally, economic theory helps point to policy measures that might ameliorate trade-offs (a familiar phenomenon in economics). However, the economic analysis of standards also has limitations.

The analysis largely focuses on the passing score required for an educational credential, for a given test, covering a given set of content standards. Consider the effect of a rise in the cutoff in a simple pass-fail situation, with a single undifferentiated diploma. All the theoretical models that we are familiar with predict a rise in the failure rate, along with other, more salutary, effects. This literature is silent on the magnitude of the rise in the failure rate (which is critical in comparison with the beneficial effects), but it does help distinguish between more and less compelling reasons for concern.4

Sorting Effects of Graduation Cutoffs

Consider first a simple sorting model, in which behavior and thus learning are held constant, independent of the standard. A rise in the cutoff merely relabels some students as failers who would otherwise be considered passers. There is, by assumption, no effect on learning or productivity. The aggregate income generated by the students is unchanged, but the distribution of it does change. The students who pass are now a more elite group, so their average productivity is higher. To the extent that graduates are pooled together in the eyes of employers (who may make only limited use of individual information, as economist John Bishop has long argued), their wages tend to rise. This point is well understood: Higher standards raise the value of a high school diploma.

Less widely understood, however, is that higher standards also raise the average quality of the pool of nongraduates, insofar as some students who would previously have passed now fail. Because nongraduates (like graduates) are evaluated by employers in part on the average quality of their pool, their wages also tend to rise. This is not a minor point. The reason nongradu-
ates typically fare so poorly under the existing system is that the ease of social promotion exacerbates the stigma attached to nongraduation.\(^5\) Thus, it is a logical fallacy to argue, as many do, that higher standards will reduce more students to the current economic level of nongraduates; the stigma on nongraduates depends on their average quality, and that depends critically on the standard itself.

A rise in standards thus leads to gains for two of the three groups—those at the top, who graduate, and those at the bottom, who would not have graduated anyway. The losers are those in the middle, who would have graduated under a less stringent standard, but who now fail. Those individuals suffer from being pooled with a group that includes those less skilled than themselves (those without the diploma) instead of with those more skilled than themselves. No efficiency loss has occurred in this pure sorting model, only a distributional effect stemming from the individuals’ relabeling. Do these losses constitute a compelling case against higher standards? The answer is no, for two reasons.

First, in terms of the narrow choice between high and low cutoffs, a high cutoff does not necessarily lead to less egalitarian outcomes. The redistribution is from the losers in the middle to the winners at both the top and the bottom. Those with the most egalitarian preferences (so-called Rawlsians, after the philosopher John Rawls) place the highest priority on raising incomes at the bottom, so they should favor a rise in standards.\(^6\) The equity implications of higher standards are not limited to those who are at increased risk of failing but include those who would fail in any case, and whose stigma stands to be reduced.

Second, the standards themselves are not at the heart of the losses from adverse pooling. The crux of the matter is the imperfect information that underlies such pooling. How disturbing is it if someone loses from no longer being confused with those of greater skill? While those able students who are now pooled with those of lesser talent should be of concern, the answer is not necessarily to reverse the rise in standards and reclassify them with those of greater talent. Perhaps, instead, information flows should be improved, if possible, such that individual talents are more accurately conveyed than with a simple binary pass-fail credential, as Bishop argued.

**Incentive Effects of Graduation Cutoffs**

The losses incurred from sorting may not be of first-order policy importance, but neither are the gains from sorting the reason for implementing
standards. The rationale for standards is to alter incentives of students, parents, teachers, and administrators to change behavior in a way that advances learning. Microeconomic analysis, the study of how rational actors respond to incentives, may offer some insights.

Economic theory predicts that the effect of raising the graduation cutoff depends on where students lie in the distribution of ability and attitudes toward study.\(^7\) Suppose the cutoff is raised from a level at which 10 percent fail to one at which 20 percent would fail under existing behavior. Under a pure sorting model, in which behavior is held constant, this rise in standards would lead to a doubling of the failure rate. Under a more realistic model, students (and their parents) respond to higher standards by reevaluating the costs and benefits of student effort.\(^8\)

How are the incentives for student effort affected at different parts of the distribution? Consider first those students at or near the twentieth percentile under the original distribution of achievement. In this example, these are students who passed under the old standard by a margin of ten percentiles, but who are now just on the margin of passing under the new standard. It would take only a small increase in their effort for a number of them to pass. The cost in doing so would be less than the substantial benefit of passing instead of failing, and so the higher standard will have a positive incentive effect on utility-maximizing individuals in this part of the distribution. As a result, one can predict with some confidence that the failure rate will not rise as much as would be naively predicted under the preexisting distribution of student achievement, because students in this part of the distribution will rise to the challenge.\(^9\)

These students, the ones for whom the most positive response is predicted and who have the most to gain from higher standards, are not the elite (they are near the twentieth percentile in this example). Unlike the elite, who will easily pass the higher standard with unchanged effort, these students are stimulated to higher effort because otherwise they will fail. These students are typically non-college-bound or marginally college-bound. For those non-college-bound students who rise to the challenge, the benefit is a high school diploma of enhanced value—a matter of great importance for those who will not have a college degree with which to distinguish themselves. For the marginally college-bound, the benefit of being prodded to meet a higher standard is better preparation for college, which, in turn, raises the probability of successful college completion.\(^10\)

However, the incentives are different farther down the distribution. Specifically, consider those students who are on the margin of failing under the old
standard (students at or slightly above the tenth percentile in this example). The effort they are exerting yields expected benefits that barely exceed the costs of the effort. A rise in the standard reduces the probability of passing with that level of effort and thereby reduces the expected benefit below the cost. For these students, the rise in standards has a negative incentive effect, leading them to reduce their effort, discouraged by the low prospects of success. They may drop out of school, as critics of standards-based reform warn. This effect is more troubling than the sorting effect, because it reduces the amount of learning in this portion of the distribution.\textsuperscript{11}

Thus, standards have different effects on students in different parts of the distribution, even among those of lesser achievement.\textsuperscript{12} Four groups of students are at risk of failing under the higher standard (see figure 1):\textsuperscript{13}

1. Some students who met previously low expectations will be stimulated to greater effort by a rise in standards, with the help of teachers and parents. (In figure 1, the dashed distribution of productivity depicts a rightward shift from just left of the new standard.) These are the most important gains from high standards.

2. Other students who would have passed under low standards will not change their behavior and will now fail. (In figure 1, these students remain
between the old and new standards, on the dashed distribution.) These stu-
dents lose, but only by virtue of being relabeled.

3. Other students, farther down the distribution, will be discouraged and
reduce effort or drop out. (In figure 1, the dashed distribution of productivity
depicts a leftward shift from just right of the old standard.) These are the most
important potential losses from high standards, toward which mitigating poli-
cies should be aimed.

4. For those students at the very bottom (the left-most portion of figure 1),
who would not pass anyway, behavior is unaffected, but they may passively
gain from the sorting effect.

Policymakers and others may differ on how to weigh the fortunes of these
groups in arriving at the optimal set of standards. The way out of this dilemma
is not necessarily to forgo the benefits of higher standards, but, if at all possi-
ble, to craft accompanying policies for those students whose efforts may flag,
especially those who might drop out.

Much of the most vocal opposition to standards-based reform, however,
comes from a completely different segment of the population—that of gen-
erally high achievers. For example, according to recent reports, “Wisconsin
scuttled plans for a high school exit exam after a protest lodged mainly by
more-affluent parents.” Similarly, efforts in Massachusetts to boycott the
statewide exams have been concentrated in affluent and high-achieving sub-
urbs, as well as high-spending communities such as Cambridge, rather than
such urban areas as Boston. State representative Ruth Balser told a group of
Brookline test critics that most of her legislative colleagues support the exams.
“It’s just those of us from districts that were already doing really well, like
Lincoln-Sudbury, Brookline, and Newton, who feel that our systems are at
risk of being dragged down by ed reform,” she said.

Perhaps the most plausible claim that suburban critics have to offer is that
higher-order skills may be deemphasized by teachers of high-achieving stu-
dents, students who are at relatively low risk of failing. It is not entirely clear
why this would be so at the high school level, if students are sorted among
basic and honors classes. The more elite students, aiming for selective col-
lege admissions, are more likely focused on the Scholastic Assessment Test
(SAT), Advanced Placement (AP) exams, and a high school transcript enhanced
with high grades in honors courses than on high school exit exams. However,
if the school reallocates resources, or changes its teaching methods to bring
up those at risk of failing, these equity-enhancing efforts could adversely
affect those of high achievement. If so, these objections to standards-based reform are not based on equity concerns, but the opposite.

The policy implication is not necessarily to forgo the benefits of higher standards, just because they may be concentrated among those for whom expectations are low, relative to the high-achieving critics. The challenge is to meet these objections by accompanying the standards with policies addressed toward the high achievers as well. This is an easier and less pressing challenge than the one concerning lower achievers, who might be discouraged from continuing academic effort.

Centralized versus Decentralized Standards

What is the proper locus of standard-setting—federal, state, or local? The movement toward standard-setting began with the states in the late 1970s (minimum competency testing), shifted toward the federal level from the late 1980s to the early 1990s, and has shifted back to the states since the mid-1990s, where it has made its greatest strides. Leaving aside the question of where content standards should be set, economic theory does have something to say about whether graduation cutoffs should be set locally or centrally.

In the simplest case, where all districts are alike, decentralization probably would lead to inefficiently low standards. Suppose each district’s non-college-bound graduates are pooled to some extent with graduates of other districts in the labor market. That is, employers do not fully distinguish graduates of any district that chooses a different standard. The reward to raising standards in any given district is thus attenuated. The district’s graduates would be of higher quality, but would not be fully identified as such, and so would only reap some of the benefits; the rest of the gains would spill over to graduates of other districts, with whom they are pooled in the labor market. As a result of this externality, local standard-setters have an incentive to free ride on the standards of other districts, establishing cutoffs that are too low to maximize their collective welfare. A centralized standard-setter would avoid this problem.

Even in this simple case, with identical districts, there are winners and losers in the choice between decentralized and centralized standards. Given that centralization raises standards, the winners are those who rise to the challenge, and the losers are those who become discouraged from exerting effort. But each district would, on the whole, be better off with a centralized standard-setter choosing the same cutoff for all districts. This logic is independent of the weights attached to winners and losers. Even the most egalitarian collec-
tion of standard-setters would prefer standards set centrally, rather than each of them riding free in a standard-cutting race to the bottom.\textsuperscript{24}

Heterogeneity across districts makes things more complicated but is also an important factor in understanding current controversies.\textsuperscript{25} For example, centralization typically raises standards in low-achieving districts but may lower them in high-achieving ones. To the extent that diplomas reflect some degree of district reputation (that is, pooling is not total), this means low-achieving districts’ graduates benefit from the rise in their standard while those from high-achieving districts lose from the drop in theirs.\textsuperscript{26} Thus, a conflict of interest may arise between those high-striving urban black students whose diploma is enhanced in value and those suburban students whose diploma could be depreciated from that obtained under decentralized standards.

With heterogeneity across districts, centralization need not always outperform decentralization.\textsuperscript{27} However, suppose the centralized standard serves as a minimum requirement for graduation, with the localities retaining the option of setting a higher standard. This arrangement outperforms decentralized standard-setting and is at least as good as central standard-setting without the local option. The result is the best of both worlds, with the centralized minimum standard putting a floor on free-riding by districts, while the high-achieving districts retain the option of exceeding that standard, if enough of the benefits accrue to their own graduates.\textsuperscript{28}

The model considered here helps frame questions that arise from current controversies. For example, in Massachusetts (among other states), the demand for local control of graduation requirements is strongest in the suburbs, while urban superintendents are generally the biggest supporters of rigorous state standards (even though their students are most at risk of failing). The urban districts suffer from a poor reputation but have still found it difficult to unilaterally raise it. One possible explanation that goes beyond the simple model but is consistent with its spirit is that a district’s reputation adjusts only slowly to its own actions. A long period of low standards will result in a low reputation, but a unilateral rise in standards may only raise the reputation over time, increasing dropouts in the short run with no reward. On this view, the imprimatur of state standards promises to be a more informationally powerful signal, more readily recognized, than the urban districts could establish on their own. Political considerations beyond the model probably are also important. The state mandate provides valuable cover to superintendents who would like to raise standards but who face local political and union obstacles to doing so and to taking steps necessary to meet them.
The model assumes some pooling, or blurring of credentials across districts even in the long run. If there is no such blurring of credentials—if each district’s diploma is fully understood by employers to represent that district’s own graduation cutoff—then the model’s case for decentralized standard-setting is stronger. But even then, high-striving students in low-achieving districts suffer from having their accomplishments depreciated by the low standards that local authorities tend to set in those districts. If policymakers are able to reduce the degree of cross-district pooling to lessen the need for centralization, then why not reduce intradistrict pooling as well, so that high achievers in any district can be evaluated by their individual accomplishments?

**Binary Credentials versus Fuller Information**

John Bishop has long argued that credentials such as a high school diploma, which convey only a binary signal to employers, are far inferior to richer and more finely graded information flows, such as those conveyed in high school transcripts. Economic theory speaks to the incentive and equity implications of improved information flows and largely bears out Bishop’s argument. A difficult question, however, is why employers often choose not to use the fuller information flows that are available. This question, to which no satisfactory answer has been found, is important in designing policies to ameliorate the trade-offs carried by a system of binary credentials.

In understanding the effects of improving information flows over that of binary credentials, sorting effects must be distinguished from incentive effects. Consider the simplest case, in which a single measure of productivity (such as a test score) is available, but a credential truncates that measure into a pass-fail signal. In a simple sorting model, in which behavior is assumed constant, the truncation of full information redistributes income by pooling. Among those who fall below the cutoff, the average income is unchanged, but it is redistributed from those just below the cutoff toward those at the very bottom, with whom they are pooled. Similarly, among those above the cutoff, the truncation of full information redistributes from those at the very top downward to those just above the cutoff. Thus, in the simplest sorting model, binary credentials generate outcomes that are more egalitarian than full information. However, even within the confines of these assumptions, the case for redistribution by blurring of differences is not compelling, unlike a case based on improved incentives.

Even before considering incentive effects, however, another aspect of sorting bears examination—the issue of job matching. Better sorting improves the
match between workers and jobs. Truncating information with a binary credential reduces the efficiency of the match and reduces output. Who bears the brunt of the lost efficiency: those at the top or those at the bottom? In one recent model the answer depends on where in the job ladder accurate sorting is most important. Suppose it is most important at the top; that is, getting the very best people into the very top jobs is more important than getting the least productive people into the very bottom jobs. Then the burden of the efficiency loss from truncating information will tend to fall on the least skilled, and this can outweigh any beneficial pooling effect they may enjoy. The wage earned by the least skilled depends on the ability of those higher up the job ladder who can only do those top jobs with the support of those lower down. If those who will fill the top jobs are not as well identified, because of truncated information, then the reward to the least skilled for supporting those in the top jobs will fall. In this case, the use of full information enhances both efficiency and equity.

Now consider the incentive effects of full information. If employers have and use individual information, diplomas and standards become irrelevant, because they add nothing to it. Each student chooses his or her own preferred level of achievement and is rewarded accordingly. More realistically, information flows can be improved by generating a discrete number of differentiated credentials. Either way, fuller information affects incentives in different ways across the spectrum of students.

Compared with a coarse pass-fail signal, better information about high achievement is a stimulus to those at the top of the distribution, who would otherwise find no payoff in exceeding the cutoff. This provides much of the answer to the criticism that high-achieving districts are dragged down by standards-based reform. High-achieving students are already motivated to excel by an array of credentials over and above high school graduation exams (for example, SAT and AP exams). If these are insufficient, differentiating diplomas based on the level of performance on the graduation exams, as a number of states do, is a relatively simple matter.

Moreover, differentiated consequences for differentiated credentials seem particularly straightforward to arrange for college-bound students. Admission to public higher education can be made contingent on higher performance levels than are required for graduation; scholarships can be based on higher levels yet. These credentials may be multidimensional, for those who find traditional graduation requirements overly narrow. For example, students can place many credentials based on artistic and musical talent on their college applications. There are literary contests, outlets such as the Concord Review (for historical essays),
and science fairs, to name just a few more credentials that high-achieving students can aim for, with confidence that they will be recognized.

Schools might arguably be under pressure to divert attention from these types of credentials toward the graduation exam, even for those students who are at no risk of failing. There could be some truth to this, insofar as districts reap rewards based on mean exam scores, instead of pass rates only (for example, the real estate market may tend to do this). However, this effect should not be exaggerated, because districts will continue to be attuned to how well their students do in college admissions, which still rests on these other types of credentials. That is why some high-achieving districts choose not to teach to the graduation exams any more than is necessary to achieve passing performance. In short, the introduction of graduation exams only adds information to the existing array of high-end credentials and should not pose any serious incentive problems for high-achieving students.

At the bottom of the distribution, the incentive effect from fuller information should also be positive. Those students who have no other way to convey their skills short of a graduation standard that is beyond their will or ability to meet would certainly gain from finer signals. As John D. Owen points out, fuller information at this end of the distribution advances egalitarian goals by giving students less extreme alternatives to dropping out.\textsuperscript{31}

This rationale is behind the proposal that students who repeatedly fail the state graduation exam might receive instead a local diploma or a local certificate of completion. Such a credential could convey the achievement of noncognitive skills such as persistence, punctuality, and discipline that are also important and rewarded in the labor market.\textsuperscript{32} The general equivalency diploma (GED) already exists as an alternative credential and should continue to signal a certain level of cognitive skills. But its payoff in the market is considerably less than a high school diploma, probably because it does not convey the same level of noncognitive skills as even a diploma based on seat time alone.\textsuperscript{33} So room remains for a credential to certify such noncognitive skills (which may be particularly important for some special education children).

The challenge is to make sure that such a noncognitive credential is properly differentiated from a standards-based credential that signifies both cognitive and noncognitive skills, and that it is treated as such by end-users (employers or colleges). This is at the heart of the dispute between those who would grant a local diploma option and those who would allow only a local certificate of completion. For reasons perhaps better understood by psychol-
ogists than economists, such terminological distinctions seem to be empirically important.

The concern is that a local diploma would not be treated with sufficient differentiation from a state diploma and would thereby undermine incentives for those students who would otherwise meet the state standard. (This seems to have been the rationale for New York’s decision to phase out the local diploma option, leaving only the Regents diploma.) A certificate of completion could and perhaps should convey the same information that a high school diploma currently conveys in those states where the requirements are almost entirely local (such as Massachusetts, until the state standards bind in 2003). Once employers recognize that a certificate of completion is equivalent to the old local diploma, there should be no basis for objecting that students are denied a diploma by the higher state standard. Diploma is only a word. If it takes a different word—certificate versus diploma—to differentiate those who have met the old local standards from those who meet the new state standards, then this would provide the finer information flows that are called for. Remaining will be those who object to such differentiation, as to all differentiation, on the grounds (perhaps unstated) that it will deny certificate-holders the benefits of being pooled with those who hold diplomas. But such sorting arguments are not persuasive.

Finally, consider students between those near the top and those near the bottom, those who would meet the state standard, but not by much. These are students for whom the incentive effects of full information are negative. They are students who rise to the challenge of the standard only because the alternatives are so much worse. If information flows are improved, these students would choose to meet a lesser level of achievement that has a lesser payoff, but not as dramatically as dropping out. The problem here is that too many students evaluate the payoffs to higher achievement differently from adults, such as their parents or state standard-setters or from the adults that they will become themselves. The labor market signals to students are somewhat remote, and many students are notoriously present-oriented. Furthermore, schools likely have a greater incentive to bring students up to a given standard when the alternative is dramatically worse than simply meeting a lesser standard. In short, while the coarse instrument of pass-fail blunts incentives for those at the bottom and the top, it does elicit greater effort from those near the passing margin.

A key policy dilemma thus emerges from our theoretical analysis: How much differentiation should exist between the state-certified standards-based
diploma and any lesser credentials? If the differentiation is too large, then students near the bottom will have no incentive to achieve beyond the low level certified by the lesser credentials. If, alternatively, the gap between the lesser credentials and the state diploma is too small (as with continuous measures, such as the test score itself, affixed to the diploma or the transcript), then too many students who might meet the state standard would be willing to settle for less, especially if employers ignore the differentiation.

Our theoretical analysis shows that some problems alleged by critics of standards-based reform are not particularly compelling, notably those based implicitly on the logic of pooling and those concerning incentives for high-achieving students. But it also points to a trade-off between incentives for those lesser achieving students who will be stimulated to meet high standards and those low-achieving students who will be discouraged. The analysis clearly indicates that the key to ameliorating this trade-off is not so much one of setting the standard high or low as it is one of filling in the information spectrum with credentials that allow lesser achieving students to demonstrate their cognitive and noncognitive skills. The optimal degree of differentiation among these credentials can probably be worked out only in practice over time, by trial and error, because it depends much on the way employers will treat different credentials, which is not something that is easily foretold.

A Description of Current State Educational Standards

Effective educational standards require the following three components:
1. Content or curriculum standards that clearly delineate what students should learn in each grade.
2. An assessment system that measures student progress toward mastery of the content standards.
3. An accountability system that stipulates a set of rewards and interventions based on student progress. Such a system should hold accountable not only students but also teachers, principals, and entire school systems for the rate of learning of students.

How close are the states to implementing educational standards that fit these criteria, and how do states vary in that regard? Complicating the analysis is that, even though standards in practice typically resemble the binary pass-fail model, these standards have taken many forms. Some states have implemented high school exit exams. Other states have left the task of assessment to indi-
A similar finding emerges from analysis of the effects of grade retention. Grade retention appears to work only when schools try to do something different, possibly with additional resources, for students as they attempt to complete a grade for a second time.

States that reduce historical inequalities in school spending before creating content standards reduce the risk of political opposition based on opportunity-to-learn lines. States that implement rigorous standards while targeting programs of demonstrated effectiveness to the students most at risk do even better.

**Conclusions and Policy Implications**

Our theoretical and empirical analysis and review of standards in practice suggests a number of conclusions and policy implications:

— Standards and accountability systems do affect incentives of students, parents, and schools. Limited, but growing, empirical evidence establishes that significant numbers of students rise to greater levels of achievement than when little was expected of them and their schools.

— Assessments should be aligned to standards; they should include open-ended questions and essays worth teaching to; and new forms should be introduced annually to avoid artificial inflation of test scores.

— Localities should retain the option to set higher standards than those set by the state.

— School financing systems should meet state constitutional requirements for adequacy or equity across districts before high-stakes standards take hold (as in Massachusetts).

— Judicious additional spending targeted at students who are likely to fail to reach standards without help makes sense. For example, programs of demonstrated effectiveness, such as Chicago’s mandatory summer school at early grade levels for those who fail to meet standards, should be replicated.

— Incentives should be strengthened for schools, especially school leaders, to ensure that students meet standards. Examples include reconstituting failing schools, reassigning teachers and administrators in these schools, and providing sanctuary for students from these schools in other schools or in new charter schools.

— Potentially harsh trade-offs can be minimized by multiple credentials, signaling different levels of achievement. Such signals already exist for high
levels of achievement. At the other end, for those students who cannot be reme-
diated to reach stipulated levels of cognitive skills, credentials need to be
developed to signal important noncognitive skills. These credentials, such as
certificates of completion, should be sufficiently differentiated from cognitive
credentials to maintain the incentive to acquire cognitive skills.

No such list of recommendations can fully anticipate what will work and
what will not work as full-blown standards-based reform takes effect. Not
everyone will meet the new standards, just as not everyone met old standards
in the past, before social promotion became the norm. New answers will
evolve to the question of what shall be done for those who fail to meet the
new standards. In the past, the GED arose to meet the needs of those who
wished to convey some level of cognitive achievement without attending
school through grade twelve. For others, alternative settings will be devel-
oped, such as the ninth-grade remedial schools in Chicago. Proposals have
been made in Massachusetts for the community colleges to admit students
into special nondegree remedial programs, for those who fail the MCAS but
receive a certificate of school completion. After-school programs analogous
to the Japanese jukus will also arise, whether by public or private initiative.

Although the optimal configuration of credentials is not yet known, one
thing is certain: It would be a disservice to all too many high school gradu-
ates to continue granting diplomas that provide no guarantee of minimal
literacy and numeracy skills. Amid all the rising controversy, it is a remark-
able fact that not even the most vocal critics of standards-based reform claim
that a diploma currently guarantees these skills. The only logical conclusion
is that those who would go back to the old system believe students should
receive a diploma even if they have not been taught basic cognitive skills, so
that they may continue to be pooled with those who have. This may seem to
be a convenient arrangement for those schools that graduate mostly high
achievers, while waving through their lagging students with a wink and a nod.
But it is no longer a credible option for those schools in disadvantaged dis-
tricts whose graduates are known to often lack basic skills and whose
communities have been notably absent from the protests against standards-
based reform.
focus time and effort on lower-achieving students. In the fourth-grade classroom that we have been studying, the teacher devoted almost a month to preparing her students for a performance assessment of the concepts of area and perimeter. The high-achieving students learned the new skills eagerly but mastered them within the first couple of lessons. They then spent the next four weeks repeating the skill over and over again while the rest of the class caught on. That same year, these children found themselves stuck with a science curriculum that never covered anything but volcanoes, largely because so much class time was devoted to test preparation. Even in districts where high-achieving and low-achieving students attend different schools, high-achieving students may still be shortchanged if administrators are forced to take resources away from high-performing schools to help low-performing schools meet the new standards.

Although we support the general principles of raising academic standards and creating a high school diploma that “means something,” we are less optimistic than Betts and Costrell about the positive consequences of standards-based reform. Assuming that children and teens will respond strongly to the positive incentives of higher standards glosses over motivational differences among children and teens who are moving through a range of developmental stages and come from a variety of ethnic and social class backgrounds, home and community support structures, and past experiences with academic success and failure. Further, placing the burden of meeting high-stakes standards on these students, without providing generous institutional support, will doom too many of them to failure. Finally, dismissing middle-class parents’ concerns as purely self-serving seems shortsighted. To maximize every child’s academic potential and maintain sufficient political support for standards-based reform, policymakers should avoid reforms that impede the academic progress of bright students from any background. In the end, standards-based reform cannot be a quick fix to the problem of low academic performance. But if policymakers consider all these complexities, it may turn out to be a step in the right direction.

Notes

1. See Public Agenda polls in recent years.
3. See Sandra Stotsky, ed., *What’s at Stake in the K-12 Standards Wars: A Primer for Educational Policy-Makers* (New York: Peter Lang Publishers, 2000). Authors such as Stephen
Arons have argued that such battles over content are a permanent feature of the public (or common) school system and can only be fully resolved by a thoroughgoing system of school choice and vouchers. Stephen Arons, *Short Road to Chaos* (Amherst: University of Massachusetts Press, 1997).

However, with or without vouchers, the demand for educational accountability in the use of public funds seems likely to rise, particularly in states where the share of funding is shifting from the localities toward the state. The specification of content standards and measurable outcomes is central to these accountability efforts.


5. A century ago, when a high school diploma was held by a small minority of the population, far less stigma was attached, economically or otherwise, to being a nongraduate. Similarly, under the traditional British system that prevailed until recently, many students left school at age sixteen. Far more students left school at this age than occurs in the United States, and the stigma was presumably much less, because their numbers included more capable workers.

6. Under this model, they should favor standards that are so high that everyone fails, so that the lowest achievers are pooled with the very best. This may seem indistinguishable from the opposite extreme, where the standard is set so low that everyone passes and is similarly pooled together. However, unless the results are perfect, with a 100 percent pass rate, the strategy of a very low standard will lead to the least egalitarian outcome, by the Rawlsian standard, because the rare failure is most highly stigmatized. In short, the wage of failers rises monotonically with the standard in this simple model. See Betts, “The Impact of Educational Standards on the Level and Distribution of Earnings.”

Robert M. Costrell relaxes a key technological assumption of this model—that the productivity of any individual is independent of other individuals (perfect substitutability, to use the technical term from economics). Suppose, instead, workers operate in teams, providing complementary services in the production of output, as in the job assignment model of Robert M. Costrell and Glenn C. Loury. Then it can be shown that another effect of raising standards works in the opposite direction from the pooling effect. High standards reduce the number of workers supported by those of lesser skill, which tends to reduce the wage of failers. Taken together with the pooling effect, raising standards need not have a monotonic effect on the wage of failers. Costrell finds that, in a benchmark case, the relationship between the wage of failers and the standard is U-shaped, and, moreover, the standard that minimizes the failers’ wage maximizes output. Costrell also analyzes the effect on this relationship of varying technology, cost of acquiring skill, and test accuracy. An important finding, however, is that those cases in which a rise in the standard reduces the wage of failers are also the cases in which equity is most likely advanced by moving away from pass-fail systems altogether, toward fuller information. Robert M. Costrell, “Are High Standards Good or Bad for Those Who Fail?” University of Massachusetts at Amherst, Department of Economics, 1999; and Robert M. Costrell and Glenn C. Loury, “Distribution of Ability and Earnings in a Job Assignment Model,” University of Massachusetts at Amherst and Boston University, 2000.


8. In addition, schools facing the prospect of higher failure rates would also respond with interventions to assist at-risk students.

9. It is an empirical matter of some importance how much less the rise would be, whether it would be closer to the full ten-point rise or closer to zero.
10. See Robert M. Costrell, “An Economic Analysis of College Admission Standards,” *Education Economics*, vol. 1, no. 3 (1993), pp. 227–41, for a formal analysis of the effect of standards in the context of college attendance, where students are uncertain how difficult college will be until they get there. A rise in admission standards forces applicants to be better prepared and can raise the resulting number of graduates, even though the number of attendees declines.

11. The analysis here excludes consideration of possible externalities created by peer effects. If there are adverse peer effects generated by some of those who are unwilling or unable to exert extra effort to pass, and if the potential benefit for some of staying in school is low, then the optimal dropout rate may not be zero. Disruptive students provide an obvious example that is unfortunately not as rare as one might hope. The best solution in such cases is not necessarily to encourage dropouts, but to create alternative educational settings for such students, such as those under creation by systems in Boston and Chicago, as long advocated by the American Federation of Teachers, among others.

12. Evidence consistent with the bifurcation in this part of the distribution is found in the contribution to this volume by John H. Bishop, Ferran Mane, Michael Bishop, and Joan Moriarty. They find that among C/C- students, minimum competency exams raise both the number of noncompleters and the number of college attendees.

13. Although the general points discussed here and depicted in figure 1 derive from the theoretical literature cited, figure 1’s continuous distribution is not strictly consistent with that literature’s simplest theoretical models. Those models generate distributions with discrete segments and a discontinuity in the vicinity of the standard.


16. It seems more likely that there could be some redistributive effect on learning in the lower grades, where heterogeneous grouping prevails.


18. This does not prevent some of the critics in these communities (both parents and educators) from couching their objections in egalitarian terms, as the defenders of those children in less-advantaged areas whose parents have chosen not to object.


21. John Bishop has provided evidence in a number of papers over the years that is consistent with this behavior of employers. See, for instance, John Bishop, “Incentives for Learning: Why American High School Students Compare So Poorly to Their Counterparts Overseas,” *Research in Labor Economics*, vol. 11 (1990), pp. 17–52.
22. The extent of this problem is inversely related to the strength of local reputation, which in turn depends on the size of the entities in question.

23. This assumes that no systematic difference exists between local and central authorities regarding the weights attached to winners and losers (that is, they hold the same social welfare function).

24. With cross-district heterogeneity, it can be the case that egalitarian societies—those that assign greatest weight to preventing dropouts—should prefer centralization even more than nonegalitarians. The problem of free-riding under decentralization is more pronounced for egalitarians because they tend to cut standards further below the optimal level. That is, egalitarians may like low standards in their own district, but they face particularly high losses from the free-riding of their fellow egalitarians in other districts choosing particularly low standards. Both egalitarians and nonegalitarians favor centralization if all districts are alike, but under cross-district heterogeneity, egalitarians may favor centralization in some cases that nonegalitarians do not.


26. Different patterns can emerge, depending on the degree of pooling. But the general point remains: There are winners and losers in any system of standard-setting, compared with any alternative.

27. It is not even certain that a centralized standard-setter would choose a higher standard than any of the localities. If the optimal central standard is tailored to the weakest districts (as it will be under some circumstances), then the central standard could end up even lower than those weak districts would choose on their own. The reason is that under decentralization, the stronger districts would choose high standards, raising the wage of non-college-bound graduates everywhere, including those in the weaker districts, to the extent they are pooled together. This would enhance the incentive for students in the weaker districts to graduate, which, in turn, allows those districts to set higher standards than otherwise without deterring too many students from graduating. In this way, it is possible that under cross-district heterogeneity central standards could be lower than under decentralization. Even if standards rise for some or all districts under centralization, the constraint that all districts face the same standard may still lead to lower social welfare than under decentralization.

28. This is the law in Massachusetts: No district will be able to award a diploma to students who fail the Massachusetts Comprehensive Assessment System (MCAS), but districts can impose additional graduation requirements, including a higher MCAS score.

29. Costrell and Loury, “Distribution of Ability and Earnings in a Job Assignment Model,” applied to the issue of standards by Costrell, “Are High Standards Good or Bad for Those Who Fail?”

30. For a formal analysis, see Costrell, “A Simple Model of Educational Standards,” section VI.


34. Economists have documented that they have a generally high rate of time preference.