10

The "First Wave" of Accountability

THOMAS S. DEE

The No Child Left Behind Act, signed into law by President George W. Bush in January 2002, mandates the development of test-based student assessments and holding schools accountable for their academic performance relative to their state's minimum threshold.[1] This bipartisan legislation will promote, and shape, the accountability systems that have been introduced in almost every state over the last several years. The accountability policies recently adopted at the state level have taken several forms, including the publication of "report cards" and ratings for schools, teacher evaluations coupled with merit pay, and the legal authority for states to control or close failing schools.[2] An increasing number of states also hold students directly accountable by withholding grade promotion or high school graduation for low performance on tests. The impetus for output-based accountability has grown out of the widely held perception that the long-standing focus of prior reforms on educational inputs and processes has been relatively unproductive. In particular, proponents of accountability policies argue that reliable information is not available on how to systematically use educational programs and resources to improve student outcomes. According to this line of reasoning, the idiosyncratic nature of educational production implies that output-based incentives provide a more reliable method for enhancing productivity.[3] However, as suggested by the title of the recent federal legislation, another clear intent of accountability programs has been to close the perfor

mance gap between advantaged and disadvantaged students, in particular the gap between white and minority students.

Critics of these policies suggest that explicit standards may exacerbate that performance gap, particularly in the absence of other systemic reforms related to local control, teacher training, and available resources.[4] Another major concern with standards-based reforms is that they may promote an undesirable narrowing of teaching styles and student curricula.[5] A recent national survey of public school teachers suggests that those concerns could be well founded. While a majority of surveyed teachers (87 percent) supported establishing higher standards, nearly 70 percent also thought that their teaching overemphasized testing to the detriment of learning in other important areas.[6] A large majority of these teachers also opposed using state tests as the sole basis for grade promotion and graduation. Nonetheless, some early indications also show that the states making the strongest recent gains in measured student achievement have been those that aggressively implemented new standards and assessments. However, the proper interpretation of the recent experiences within particular states has been the subject of considerable controversy, and many of the new state policies have been implemented too recently to be evaluated with currently available data.[7]

The premise for this paper is that useful insights into these controversies may be gained by looking back to consider the consequences of earlier state-level standards. Over the last twenty-five years, almost every state introduced stricter, state-level standards for high school graduation in response to highly publicized concerns about student effort and the quality of public schools. The key first-wave reforms consisted of a test-based performance standard, minimum competency testing (MCT), and a process standard—course graduation requirements (CGRs) that mandated the amount of academic credits that must be earned in core academic areas. Though the adoption of these reforms occurred with much fanfare, surprisingly little study has been conducted since then of their consequences. In this chapter, I describe these reforms and what is currently known about their implications for student outcomes. I also present new evidence on how these reforms influenced a variety of student outcomes (for example, educational attainment, labor market experiences, high school curricula) in empirical specifications that address the possible shortcomings of prior evaluations. I conclude by discussing what this broad set of results may contribute to our understanding of the currently evolving state and federal experiments with standards-based reform.

The First Wave of Education Reform

Much of the ongoing public interest in reforming public education can be traced to the mid-1970s and the widely discussed evidence that student test scores and the quality of public schooling were in decline.[8] In particular, critics of that period emphasized that a high school diploma, once a significant and hard-earned, personal accomplishment, had been debased through the abuses of social promotion and the tolerance of low academic standards.[9] Politicians at the state level proved highly responsive to these concerns and began enacting a variety of new standards and regulations now known as the first wave of education reform.[10] The earliest manifestation of these centralized reforms was the widespread adoption of a test-based performance standard: minimum competency testing. Beginning in 1975 nearly every state introduced new MCT programs designed to assess students' basic skills.[11] Most of these programs were simply intended to identify low-performing students and to direct them to sources of remediation. However, several of these states also mandated that students pass a minimum competency test to graduate with a standard diploma. By 1992 the graduating high school seniors in fifteen states were required to pass such a test.[12] Typically, students would first sit for these exams in the ninth or tenth grade and have multiple opportunities for retests. The conventional wisdom regarding these test-based diploma sanctions has been that they were "legislated as a lion but implemented as a lamb" (see chapter 3 in this volume).[13] Specifically, the MCT standards typically required that students demonstrate basic math and reading skills at only an eighth- or ninth-grade level. Furthermore, in response to failure rates on initial tests that were deemed politically unacceptable, these standards were sometimes lowered. As a consequence, the ultimate pass rates among high school seniors were extremely high.[14] However, whether MCT has had a more substantial influence on dropout rates is an open empirical question because the attrition of discouraged students may make the ultimate pass rates misleading. Furthermore, a full consideration of MCT policies should also consider their effects on other outcomes that are relevant to all students (for example, student curricula and labor market experiences).

The adoption of first-wave reforms accelerated more dramatically in the early 1980s after the publication of several panel reports, which were highly critical of public education. The most widely discussed of these reports, *A Nation at Risk*, emphasized the need for higher expectations and standards for high school graduates.[15] In particular, the report alleged that the combination of a "cafeteria-style curriculum" and "extensive student choice" meant

that too many students pursued a diffuse and unchallenging course of study. The report recommended that states respond with new high school graduation standards mandating a minimum amount of course taking in core academic areas. The report specifically recommended a "New Basics" curriculum requirement consisting of four years of English and three years each of social studies, science, and mathematics.[16] Again, politicians proved highly responsive to the strong public interest in these policies. By 1992 nearly every state had increased its course graduation requirements in the four core academic areas.[17] However, in all but three states (Florida, Louisiana, and Pennsylvania), the new CGRs fell short of the "4/3/3/3" standard recommended by *A Nation at Risk*.

Standards and Student Achievement

The fundamental motivation for these two first-wave reforms (MCT and CGRs) was simply to promote student effort and learning, making high school diplomas worth the paper they are written on. Are new educational standards likely to have beneficial effects on student achievement? Commentators on this issue have disagreed sharply. Proponents of higher standards make the straightforward claim that the "incentive effects" of such policies will raise the level of achievement among those students who would pass under a weak standard and choose to increase their effort to meet the new standard.[18] Julian R. Betts and Robert M. Costrell also suggest that those students whose prior levels of effort would clearly imply failing or passing both standards (that is, those at the top and bottom ends of the ability distribution) will not have any incentive to change their behavior. However, they also recognize that the incentive effects for some students who marginally passed under weaker standards will promote discouragement and reduced effort. They recommend targeted policies to attenuate these losses.

However, the potential benefits of higher standards are not necessarily limited solely to those marginal students who choose to increase their effort. For example, John H. Bishop discusses how a high, external standard can limit the "nerd harassment" and peer pressure that encourages high-ability students to shirk educational effort.[19] Standards may also generate broader educational gains through general increases in educational expectations and school productivity. Furthermore, even those students who fully anticipate dropping out of high school may be compelled in the short term to greater educational effort through curricular mandates such as course graduation requirements. The "sorting effects" of higher standards may also lead to passive labor market rewards.[20] Specifically, if educational attainment functions

as a signal of unobserved individual ability, higher standards could increase the attractiveness to employers of all students by increasing the average level of ability among both dropouts and graduates.

The critics of standards-based reform emphasize the many negative consequences associated with the expected reductions in educational attainment. Furthermore, they note that the reductions in educational attainment are disproportionately likely to be among those whose poor socioeconomic backgrounds make it unusually difficult to meet new standards. In particular, several observers suggest that higher standards will exacerbate the troublesome performance gaps between black and white students.[21] The consequences of higher standards for the racial gap in educational performance may be driven by more than simple differences in socioeconomic backgrounds. Higher standards and high-stakes testing may also harm minority students if they generate "stereotype threat": academic underperformance due to the risk of confirming negative stereotypes.[22]

However, the critics of standards also suggest that these reforms will have other, pejorative effects that harm all students. For example, the introduction of high-stakes testing such as minimum competency tests may lead to a narrowing of teaching styles and curricula (that is, "teaching to the test") that comes at the expense of substantive learning.[23] Furthermore, the establishment of minimum competency tests and stricter course graduation requirements may suggest to students that learning for its own sake is not worthwhile. In particular, these standards may encourage otherwise high-achieving students to avoid challenges and simply choose the path of least resistance to satisfying their requirements.[24] The authors of *A Nation at Risk* made a similar allegation, suggesting that minimum competency tests were inadequate because they would become maximum standards and lower expectations for high-ability students.

Evidence from the 1990 PUMS

These disagreements suggest that fundamental, policy-relevant issues about the educational consequences of higher standards can be informed only by empirical evidence. Given this, one might expect that the first-wave reforms have been subjected to exhaustive empirical evaluation. Surprisingly, relatively little empirical evidence exists about the consequences of these policies that would allow for sorting through these conflicting theoretical predictions. Furthermore, what evidence is available is often directly contradictory.[25] One possible explanation for these conflicting results is that almost all of the prior empirical studies have effectively relied on cross-sectional comparisons of stu-

dents who reside in states with different policies.[26] Educational outcomes vary considerably across states, reflecting a variety of cultural, socioeconomic, and political determinants that are often inherently difficult for researchers to measure directly. These unobserved but state-specific determinants of educational achievement are also likely to be associated with each state's propensity to adopt high school graduation requirements such as MCT and CGRs. This implies that the results of cross-state comparisons may be sensitive to the presence of additional control variables and subject to biases of an unknown direction.[27] A second drawback of prior empirical studies is that they have not directly addressed claims about whether these graduation standards would be particularly harmful or beneficial to minority students. In this section, I present new evidence that addresses both of these concerns by relying on individual-level data from the 1990 Public Use Micro-Data Sample (PUMS).

Data and Specifications

The 1990 PUMS consists of approximately twelve million respondents (5 percent of the population) who completed the long-form questionnaire to the decennial census. One useful feature of the PUMS is that the large number of PUMS respondents implies increased statistical precision and, in particular, a better ability to detect race-specific responses to the new graduation standards. My extract from the PUMS data consists of the 1,348,766 white (non-Hispanic) and black respondents who were age eighteen between 1980 and 1988 and born in one of forty-nine states.[28] Two of the outcome variables defined for each respondent identify educational attainment, a binary indicator for high school graduation (mean = 0.858) and another for college entrance (mean = 0.519).[29] I limited the sample to those who were at least eighteen by 1988 because of the biases that could be generated by state-specific trends in the "incomplete spells" of high school completion and college entrance among cohorts who were younger at the time of the census interview.[30] The other dependent variables reflect the labor experiences of each PUMS respondent. One is a binary indicator for employment participation (mean = 0.745), which is defined for all respondents.[31] The other is the natural log of average weekly wages, which is only defined for 1,143,352 respondents. This wage variable is the ratio of pre-tax wage and salary income reported for the previous calendar year and the corresponding number of weeks worked (mean = 327).

Another particularly useful feature of the PUMS is that it provides data from different birth cohorts within each state. This allows me to construct before-and-after comparisons of students within the same states, instead of relying exclusively on cross-state comparisons. My research method effec-

tively begins by comparing individuals of different ages within each state, some of whom attended high school before the first-wave education reforms were implemented and others who attended afterward—and were thus required to take minimum competency exams and more courses in academic areas to graduate. To eliminate the influence of age and other national trends, my methodology also compares the changes within states that introduced first-wave reforms (treatment states) with the contemporaneous changes in states that did not (control states). More specifically, these types of comparisons are effectively made by relying on estimates from the following multiple-regression model:

$$Y_{ist} = \beta X_{ist} + \gamma Z_{st} + \mu_s + \nu_t + \varepsilon_{ist}, \tag{1}$$

where Y_{ist} is the dependent variable and the matrix, X_{ist}, includes observed, individual-level traits. In most models, these controls simply include binary indicators for race and gender. However, in the models for labor market outcomes, these controls also include measures of educational attainment (that is, separate dummy variables for high school graduates, those with some college, and those with bachelor's degrees) and a dummy variable for whether the respondent attended school within the last year.[32] The terms μ_s and ν_t represent fixed effects specific to each state of birth and year of birth. The term ε_{ist} is a mean-zero random error.[33]

The matrix, Z_{st}, includes the observed determinants that were specific to the birth cohorts within each state. These variables include the two independent variables of interest: dummy variables that reflect the state MCT and CGR policies in effect for each birth cohort at age eighteen. One dummy variable simply indicates whether a minimum competency test was required for that particular graduating class. The second dummy identifies whether the state had a high, academically focused CGR in effect for that graduating class. A high CGR is defined here as a required high school curriculum that includes at least three Carnegie units in English, two in social studies, one in science, and one in mathematics.[34] These and other state-year controls were matched to the respondents by their state of birth and year of birth. One noteworthy limitation of the PUMS data is that relying on state of birth may introduce measurement error because some children have moved to different states by the time they reach high school. However, the attenuation bias implied by such measurement error suggests that the reported estimates can be interpreted as lower bounds on the true effects.[35]

The identification strategy embedded in this model makes a potentially important contribution to the understanding of the consequences of first-

wave reforms because it removes the possible biases resulting from unobserved state-level determinants. The model effectively does this by comparing the cohort differences in the treatment states before and after the introduction of new standards to the contemporaneous cross-cohort changes in the control states. I present some evidence on the empirical relevance of relying on within-state versus cross-state comparisons by comparing the results of models that do and do not include the state fixed effects, μ_s. I also present some heuristic evidence on this specification issue through the use of a simple counterfactual in which I estimate the effect of a state policy that should not have large and statistically significant effects on educational attainment. To the extent that a particular specification suggests that this policy did have large and statistically significant effects, the existence of specification error is suggested.[36]

In the preferred specifications, which include state fixed effects, the possible sources of omitted variable biases are the unobserved determinants of Y that are also related to the timing of new standards within states. The matrix, Z_{st}, addresses this concern by including other regression controls that vary by state and year. For example, new state standards were sometimes part of omnibus education bills that included other policy changes such as increased spending. To control for the possible effects of school spending, some models include, as an independent state-level variable, real expenditures on K–12 public schools per student in average daily attendance when the respondents were sixteen to seventeen years old. For example, respondents who were eighteen in 1980 were matched to the school expenditures in their state during the 1978–79 school year. Another state-year control in most models is the state unemployment rate when the respondent was seventeen years old. This variable is expected to have a positive effect on educational attainment because it reduces the opportunity costs associated with remaining in school.[37] David Card and Thomas Lemieux present evidence that the natural variation in the size of a particular birth cohort's population can also influence educational attainment.[38] At the college level, this could occur if temporary increases in cohort size are not fully matched by an increased supply of enrollment space at local colleges and universities. At the secondary level, increased cohort size may reduce the benefits of remaining in school by lowering school quality.[39] Therefore I also include a measure of cohort size based on the natural log of the U.S. Census Bureau's estimate of eighteen-year-olds in the respondent's state of birth at age eighteen. I also include a measure of the real costs of postsecondary tuition based on the in-state rate at "lower-level" state colleges and universities when the respondent was seventeen years old.[40] Finally, as a control for within-state changes in socioeconomic back-

ground, I also matched all respondents to the poverty rate in their states when they were seventeen years old.

Results

In table 10-1, I present the estimated effects of MCT and CGR policies on educational attainment across a variety of specifications. These results demonstrate that the estimated effects of the first-wave reforms on educational attainment are sensitive to controlling for unobserved state fixed effects. For example, the models that exclude state fixed effects but include the other state-year controls suggest that MCT significantly reduced the probability of graduating from high school and attending college. These results also suggest that CGR policies had no statistically significant effects on either measure of educational attainment. However, the models that include state fixed effects and the other state-year controls imply that MCT had small and statistically insignificant effects on both outcomes. These models also suggest that higher CGRs reduced the probability of graduating from high school by a statistically significant 0.48 percentage points. An effect of this size represents a 0.6 percent reduction in the mean probability of graduating from high school, or, alternatively, a 3 percent increase in the mean probability of dropping out. Another way to frame the size of this estimate is to note that high school completion rates among eighteen- to twenty-four-year-olds increased from 82.8 percent in 1972 to 86.5 percent in 2000. My results suggest that in states that adopted high curricular standards these average gains were attenuated by approximately 14 percent.

Several dimensions of the results in table 10-1 suggest that the inferences from the models with state fixed effects are more reliable. First, F-tests indicate that the state fixed effects are jointly significant determinants of educational attainment. Second, in models that exclude state fixed effects, the key results are highly sensitive to the presence of the other state-year controls. For example, similar to prior studies, the first model implies that MCT significantly increased the probability of attending college. However, after introducing the other controls, this estimate becomes negative and significant. This type of sensitivity suggests the difficulty of relying on proxies for the determinants of educational achievement across states. Third, the sensitivity of these evaluation results to the introduction of state fixed effects does not appear to reflect any loss of sampling variation or statistical precision. Specifically, some of the cross-state models suggest that MCT and CGRs had effects roughly 2 percentage points in size. However, in the preferred specification (column 5 of table 10-1), the standard errors are sufficiently small that effects of that size can be rejected at conventional levels of statistical significance.

Table 10-1. *Linear Probability Models for Educational Attainment*

Independent variable	Estimated effects					
High school graduate						
Minimum competency test	0.0001	−0.0187***	−0.0165***	−0.0013	−0.0014	−0.0015
	(0.0065)	(0.0033)	(0.0031)	(0.0020)	(0.0020)	(0.0020)
High course graduation requirement	−0.0051	−0.0027	−0.0020	−0.0055**	−0.0048**	−0.0048**
	(0.0041)	(0.0024)	(0.0023)	(0.0025)	(0.0022)	(0.0022)
Any state executions at age eighteen?	...	...	−0.0126***	...	...	0.0004
			(0.0041)			(0.0019)
College entrant						
Minimum competency test	0.0244**	−0.0240***	−0.0211***	−0.0054*	0.0002	0.0009
	(0.0113)	(0.0056)	(0.0054)	(0.0028)	(0.0030)	(0.0030)
High course graduation requirement	−0.0197***	−0.0058	−0.0047	−0.0079*	−0.0028	−0.0028
	(0.0075)	(0.0046)	(0.0045)	(0.0048)	(0.0033)	(0.0033)
Any state executions at age eighteen?	...	...	−0.0170***	...	...	−0.0026
			(0.0064)			(0.0024)
State fixed effects?	No	No	No	Yes	Yes	Yes
State-year controls?	No	Yes	Yes	No	Yes	Yes

Source: Author's calculations based on the 1990 Public Use Micro-Data Sample (PUMS), state-year controls for state graduation requirements, the unemployment rate, the natural log of cohort size, the poverty rate, real K–12 expenditures per pupil, and real postsecondary tuition.

Note: The PUMS extract consists of the 1,348,766 white (non-Hispanic) and black respondents who were born in one of forty-nine states (Nebraska excluded) and who were age eighteen between 1980 and 1988. All the models include fixed effects for race, gender, and year of birth. Heteroscedastic-consistent standard errors, adjusted for state-of-birth by year-of-birth clustering, are reported in parentheses.

* Statistically significant at 10 percent level.
** Statistically significant at 5 percent level.
*** Statistically significant at 1 percent level.

A final way to provide some ad-hoc evidence on the reliability of cross-state versus within-state comparisons is through the use of a simple counterfactual. To the extent that empirical evaluations relying on cross-state comparisons generate reliable results, conditional on the other controls, irrelevant state policies should have small and statistically insignificant effects on educational attainment. However, to the extent that an irrelevant policy appears to have a large and statistically significant effect, it suggests the existence of biases driven by the unobserved, state-specific determinants of educational outcomes. The results in table 10-1 present such evidence by reporting the estimated effects on educational attainment of having any state executions at age eighteen. The models without state fixed effects suggest that capital punishment generates large and statistically significant reductions in the probability of high school completion (1.3 percentage points) and college entrance (1.7 percentage points).[41] However, in the models that rely on the within-state variation in executions, these estimates are much smaller, more precisely estimated, and statistically insignificant.

In table 10-2, I present the key evaluation results from the preferred specifications that include both the state and year fixed effects and the state-year controls. I also report the estimated effects of these first-wave reforms from separate models for white males, white females, black males, and black females. All of these models suggest that the first-wave reforms had statistically insignificant effects on the probability of entering college. The absence of any effects on college entrance is plausible because these high school graduation requirements are less likely to be binding for the relatively high-achieving students on the margin for attending college. However, the results also indicate that these reforms had fairly large and statistically significant effects on the probability of completing high school and that these effects varied considerably by race and gender. In particular, the estimates suggest that higher CGRs significantly reduced the probability of completing high school for white males and blacks but not for white females.[42] Notably, the reform-driven reductions in educational attainment were particularly large among blacks (roughly 2 percentage points). These estimated reductions are roughly four times larger than those for white males. Similarly, the results suggest that the only large and statistically significant effect of introducing MCT was among black males who experienced an estimated 1.26 percentage point reduction in the probability of completing high school. The estimated effects of these first-wave reforms are also fairly large relative to the recent growth in educational attainment among blacks. Between 1972 and 2000, the high school completion rate of eighteen- to twenty-four-year-old blacks increased from 72.1 percent to 83.7 percent.

Table 10-2. *Estimated Effects of Minimum Competency Tests (MCT) and High Course Graduation Requirements (CGR) on Educational Attainment by Race and Gender*

| | Dependent variable | | | | |
| | High school graduate | | College entrant | | |
Sample	MCT	CGR	MCT	CGR	Sample size
All respondents	−0.0014	−0.0048**	0.0002	−0.0028	1,348,766
	(0.0020)	(0.0022)	(0.0030)	(0.0033)	
White males	0.0025	−0.0053**	0.0049	−0.0026	585,376
	(0.0033)	(0.0025)	(0.0045)	(0.0040)	
White females	−0.0022	−0.0005	0.0007	−0.0017	588,611
	(0.0026)	(0.0032)	(0.0035)	(0.0035)	
Black males	−0.0126**	−0.0211**	−0.0024	−0.0161	81,799
	(0.0055)	(0.0087)	(0.0073)	(0.0111)	
Black females	0.0010	−0.0203***	0.0044	−0.0188	92,980
	(0.0054)	(0.0070)	(0.0070)	(0.0015)	

Source: Author's calculations based on the 1990 Public Use Micro-Data Sample (PUMS), state-year controls for state graduation requirements, the unemployment rate, the natural log of cohort size, the poverty rate, real K–12 expenditures per pupil, and real postsecondary tuition.

Note: The PUMS extract consists of the 1,348,766 white (non-Hispanic) and black respondents who were born in one of forty-nine states (Nebraska excluded) and who were age eighteen between 1980 and 1988. All the models include the state-year controls and fixed effects for race, gender, state of birth, and year of birth. Heteroscedastic-consistent standard errors, adjusted for state-of-birth by year-of-birth clustering, are reported in parentheses.

** Statistically significant at 5 percent level.

*** Statistically significant at 1 percent level.

The evidence from table 10-2 is largely consistent with the concerns sometimes raised by critics of standards-based reforms.[43] The introduction of high school graduation standards led to reductions in educational attainment that were particularly concentrated among black students. These effects could stem from a race-specific phenomenon such as stereotype threat. Regardless, these results suggest that the largest impact of higher standards will be upon those students whose socioeconomic background puts them at high risk for academic failure. However, a full evaluation should also consider the implications of these reforms for labor market experiences. Attention to the labor market consequences of these policies also has a strong intuitive appeal given that local business leaders concerned with the quality of their work force were often instrumental in the adoption of first-wave reforms. Higher standards may benefit students (even those who drop out) by inducing increased educational effort that is rewarded in the labor market (an incentive effect). There may also be distributional consequences of these

Table 10-3. *Estimated Effects of Minimum Competency Tests (MCT) and High Course Graduation Requirements (CGR) on Employment Participation and Wages by Race and Gender*

| | Dependent variable | | | |
| | Employed | | Log wages | |
Sample	MCT	CGR	MCT	CGR
All respondents	−0.0053*	0.0081**	0.0088	−0.0074
	(0.0031)	(0.0035)	(0.0060)	(0.0070)
White males	0.0012	0.0106***	0.0094	−0.0075
	(0.0033)	(0.0041)	(0.0069)	(0.0078)
White females	−0.0095**	0.0011	−0.0073	−0.0041
	(0.0031)	(0.0044)	(0.0066)	(0.0070)
Black males	0.0164**	0.0339***	0.0108	−0.0279
	(0.0069)	(0.0101)	(0.0148)	(0.0177)
Black females	−0.0025	0.0182*	0.0169	0.0251
	(0.0080)	(0.0104)	(0.0133)	(0.0166)

Source: Author's calculations based on the 1990 Public Use Micro-Data Sample (PUMS), state-year controls for state graduation requirements, the unemployment rate, the natural log of cohort size, the poverty rate, real K-12 expenditures per pupil, and real postsecondary tuition.

Note: The PUMS extract consists of the 1,348,766 white (non-Hispanic) and black respondents who were born in one of forty-nine states (Nebraska excluded) and who were age eighteen between 1980 and 1988. Log wages are only defined for 1,143,352 respondents. All the models include the state-year controls, individual-level controls for educational attainment, and student status and fixed effects for race, gender, state of birth, and year of birth. Heteroscedastic-consistent standard errors, adjusted for state-of-birth by year-of-birth clustering, are reported in parentheses.

* Statistically significant at 10 percent level.

** Statistically significant at 5 percent level.

*** Statistically significant at 1 percent level.

reforms to the extent that higher standards increase the prestige of being a high school graduate and correspondingly reduce the stigma associated with being a dropout (a sorting effect).

Table 10-3 presents new evidence on these issues by reporting the estimated effects of the first-wave reforms on employment participation and log wages for the full PUMS sample and for samples defined by race and gender. These models include state and year fixed effects, the state-year controls, and additional individual-level controls for educational attainment and student status. Unlike the prior cross-sectional evaluations, these results suggest that both reforms had small and statistically insignificant effects on wages for all groups. However, the results also suggest that the first-wave reforms had some statistically significant effects on employment participation. For example, a minimum competency test significantly reduced the probability of

employment for white females by 0.95 percentage points.[44] But most of the reform-driven changes in employment were positive. The existence of employment gains and the simultaneous absence of significant wage effects indicate that the labor market gains may largely reflect improved signaling and not productivity gains among students. The reform-driven changes in employment also appear to vary considerably by race. For example, the estimates in table 10-3 indicate that higher CGRs increased the probability of being employed by roughly 1 percentage point for white males and by 3 percentage points for black males. The introduction of MCT also increased the probability of employment for black males by a statistically significant 1.64 percentage points but had a smaller and statistically insignificant effect among white males. One useful way to underscore the magnitude of these race-specific policy effects is to note that, in these data, white males were roughly 19 percentage points more likely to be employed than black males. Because the employment gains attributable to each first-wave reform were roughly 2 percentage points larger for black males than for white males, states that implemented one of them closed the black-white employment gap by roughly 11 percent.

The results in tables 10-2 and 10-3 suggest that new CGRs were a meaningfully binding standard that had educational and labor market consequences for almost all students. In contrast, the effects of MCT were more limited. These results are consistent with the anecdotal evidence suggesting that minimum competency tests were "implemented as a lamb" in response to political realities. The results also suggest that, when binding, higher standards of either type had decidedly mixed distributional consequences. They reduced educational attainment, particularly among black students. However, they also generated some labor market rewards in the form of increased employment probabilities that were also concentrated among black students. How can these gains and losses be compared? One possibly useful point of reference is the expected wage associated with being a high school graduate or a dropout. A rough calculation based on these data suggests that high school graduates receive an expected wage premium equal to approximately 33 percent of a dropout's average wage, w_d.[45] This implies that those who dropped out of school in response to the higher standards suffered substantive consequences. Their loss of this wage premium was offset only somewhat by a 0.0081 increase in the probability of employment as a dropout. However, for those who would have dropped out or graduated without regard to the changed graduation requirements, there were unambiguous labor market gains because they were significantly more likely to be employed.

Another possible useful way to frame these costs and benefits is to ask how

these reforms might change the expected wage for someone who was uncertain about whether he or she would be a high school graduate or not. For such a person, the expected cost of higher CGRs is related to the reduced probability of enjoying the 33 percent wage premium of high school graduates. This expected cost equals $0.0016w_d$ (that is, 0.0048 x 0.33). The expected benefit of a higher CGR is related to the increased probability of being employed (0.0081) at an expected wage equal to $1.155w_d$.[46] This expected benefit equals $0.0094w_d$. Therefore, the expected wage benefits of a higher CGR exceed expected wage costs by a factor of roughly 6 (0.0094/0.0016).[47] This suggests that a risk-neutral person might prefer a regime with higher standards to one without and that the net effects of the higher standards on expected wages are positive. However, these back-of-the-envelope calculations do not constitute a full cost-benefit analysis. In particular, the comparisons ignore distributional consequences as well as the other social losses that may be associated with reform-induced reductions in educational attainment (for example, those related to health, criminal, and civic behaviors).

Standards and Educational Processes

The evidence suggests that first-wave reforms sometimes reduced educational attainment and also generated some improvements in the probability of employment. These labor market consequences of stricter graduation standards could, in most cases, simply reflect passive sorting effects. However, they could also indicate some reform-induced increases in educational effort, which were subsequently rewarded in the labor market (that is, incentive effects). In this section, I provide some empirical evidence on the second possibility by examining how the first-wave reforms influenced several educational process measures: the amount of academic course taking among individual high school students. Academic course taking among public high school graduates did increase significantly during the eighties across students of varying demographic traits.[48] These increases were particularly large in mathematics and science. For example, the average number of Carnegie units among public high school graduates in 1982 and 1994 increased 27 percent in mathematics and 38 percent in science.[49] Some studies have suggested that the new CGRs were at least partially responsible for these increases. For example, Stanley Legum and others find that the high school graduates in states with higher CGRs have higher levels of academic course taking.[50] Similarly, William H. Clune and Paula A. White, in a study of four states, found that academic course taking among graduates increased after the introduction of more demanding CGRs.[51] However, two specification issues may bias

these inferences about the effectiveness of first-wave reforms. One is that the appearance of policy-induced increases in academic course taking could simply be the result of the increased dropout rate (table 10-2) instead of genuine increases in academic effort. A second concern is that the identification strategies, which rely exclusively on either cross-state or time-series comparisons, may lead to substantively biased inferences.[52] The evaluations presented here provide new evidence on these issues by examining estimates from models that include eventual dropouts and that control for unobserved state and year fixed effects.

These evaluations also provide new evidence on some important concerns raised by critics of standards-based reform. Specifically, they suggest whether higher educational standards might have unintended and pejorative effects on educational processes. For example, the models for academic course taking presented here provide some evidence on whether the creation of minimum standards led to high-performing students reducing their curricular effort in the core academic areas.[53] I also discuss evidence on how such effects differed for high- and low-performing students and whether the first-wave reforms narrowed student curricula by reducing Carnegie units in the visual and performing arts. And I consider how these reforms may have influenced the intellectual engagement of students as measured by changes in time spent reading for pleasure, watching television, and doing homework.

Data and Specifications

The data for these evaluations were created by pooling observations from two major longitudinal studies fielded by the National Center for Education Statistics: the sophomore cohort from High School and Beyond (HS&B) and the National Education Longitudinal Study of 1988 (NELS). These surveys provide student-level data from before and after the time when most first-wave reforms were implemented. More specifically, HS&B and NELS provide nationally representative samples of tenth-grade students from 1980 and 1990, respectively.[54] Because each of these studies had a transcript component, they also include data on the Carnegie units earned in particular subject areas in addition to survey questions on students' use of time.[55] My extract from these surveys consists of white non-Hispanic and black respondents who were tenth graders in 1980 (HS&B) and 1990 (NELS) and includes eventual dropouts. The combined sample size with available transcript data consists of 18,134 observations (9,331 from HS&B and 8,803 from NELS).[56]

The econometric specification I used for models based on these data is similar to the preferred specification from the previous section. The independent variables of interest reflect the state high school graduation require-

ments in effect for the graduating classes of 1982 and 1992, respectively. The other independent variables include state and year fixed effects, individual-specific variables, and variables specific to each state-year cell. The individual-level controls include single dummy variables for race, gender, and age (born before 1964 for HS&B respondents and before 1974 for NELS respondents). These controls also include four dummy variables for the highest level of parental education, five dummy variables for family composition, and four dummy variables for quartiles of socioeconomic status (including one for a missing socioeconomic status index). I matched each respondent to the relevant graduation requirements and other state-year controls by exploiting the state identifiers in the restricted-use versions of these surveys.[57] The state-year controls again include 1981 and 1991 data on real public school spending per capita, the state unemployment rate, the poverty rate, and the real postsecondary tuition level. I also matched each respondent to the size of his or her state-year cohort: 1982 and 1992 data on the natural log of the eighteen-year-old population in the state.

Results

In table 10-4, I present estimates of how first-wave reforms influenced the amount of academic credit earned in the four core subject areas. These results uniformly suggest that the introduction of MCT reduced course taking in these academic areas. However, these estimated reductions are statistically significant only in the sciences and mathematics. The estimated effects in these two subjects are roughly equal to 5 percent of the dependent means. The reductions in curricular effort could reflect lowered motivation among students who clearly exceeded the testing standards or a possibly unintended reallocation of school and teacher resources toward lower performing students. But there could be an additional ambiguity to these results as presented because these effects might simply be explained by students who were induced into dropping out (table 10-2), thereby taking fewer courses. However, other evidence suggests that these effects reflect real policy-induced reductions in curricular effort among conventional students. For example, the introduction of MCT only appeared to increase the probability of dropping out among black males (table 10-2). But the reductions in course taking associated with MCT were more uniform across demographic traits and were particularly large for female students. Furthermore, the introduction of MCT also led to large and statistically significant reductions in calculus credits, a margin only relevant for high-achieving students.[58]

In contrast to the MCT results, the evidence in table 10-4 suggests that higher CGRs had the desired effect of generating substantive increases in the

Table 10-4. *Estimated Effects of Minimum Competency Tests (MCT) and High Course Graduation Requirements (CGR) on Carnegie Units by Academic Subject*

Academic subject	Dependent mean	Independent variable		
		MCT	1–2.99 units required in subject	3+ units required in subject
English	3.7	–0.067	0.032	0.329***
		(0.056)	(0.051)	(0.051)
Social studies	3.1	–0.088	–0.015	0.133
		(0.084)	(0.098)	(0.101)
Science	2.4	–0.135**	0.092*	0.393***
		(0.055)	(0.052)	(0.084)
Mathematics	2.7	–0.127**	–0.020	0.110*
		(0.050)	(0.046)	(0.059)

Source: Author's calculations based on High School and Beyond (HS&B), the National Education Longitudinal Study of 1988 (NELS), state-year controls for state graduation requirements, the unemployment rate, the natural log of cohort size, the poverty rate, real K–12 expenditures per pupil, and real postsecondary tuition.

Note: The HS&B and NELS extract consists of the 18,134 white (non-Hispanic) and black respondents (Nebraska excluded) who were in tenth grade in 1980 (HS&B) or 1990 (NELS). All the models include the state-year controls, individual-level controls for race, gender, age, parental education, family composition, socioeconomic status quartile, and state and year fixed effects. Heteroscedastic-consistent standard errors, adjusted for state-year clustering, are reported in parentheses.

* Statistically significant at 10 percent level.
** Statistically significant at 5 percent level.
*** Statistically significant at 1 percent level.

credits earned in these core areas. In particular, a high CGR (three or more Carnegie units required in a particular subject) led to increased credit in each academic subject.[59] For example, a high CGR in science increased credits earned by 0.393 relative to the reference category of a weak or nonexistent CGR (less than one Carnegie unit required in subject). This estimated effect is roughly equal to 16 percent of the mean science credits. However, the estimated effect of a high CGR on social studies was statistically insignificant as were the much smaller estimated effects of weaker CGRs (1–2.99 Carnegie units required). Nonetheless, these results suggest that new CGRs did contribute substantively to the academic upgrading of high school curricula over this period, particularly in English and the sciences. For example, the estimated effect associated with a high CGR in science is equal to roughly 60 percent of the average growth in science credits over this period.

Overall, the results in table 10-4 suggest that MCT sometimes had negative effects on curricular outcomes while the effects of higher CGRs were often positive. However, one of the difficulties of interpreting the CGR results is that they define the policies' effects for the average student. The introduction of higher CGRs could conceivably have had very different effects among high- and low-performing students. In particular, the results in table 10-4 could be misleading because the introduction of higher CGRs may have simultaneously reduced academic course taking among the low-performing students induced into dropping out. However, that does not appear to be the case. The introduction of higher CGRs appears to have increased academic course taking among those at risk for dropping out. Specifically, similarly specified models indicate that higher CGRs increased the probability of having at least one Carnegie unit in these academic subjects (particularly English and the sciences).[60] Other concerns about student-level standards involve whether they narrow student curricula in undesirable ways or reduce the intellectual engagement of students. To address the first issue, I used similarly specified models to estimate the effects of the first-wave reforms on student involvement in the visual and performing arts (as measured by Carnegie units) and on student participation in school musical activities (that is, the school band, orchestra, or chorus). For both outcomes, the estimated effects of the first-wave reforms were imprecisely estimated and statistically indistinguishable from zero. However, I found that the first-wave reforms did influence proxy measures of intellectual engagement: the amount of time students spent reading for pleasure, watching television, and doing homework. In particular, a higher CGR was associated with large and statistically significant reductions in the amount of time spent reading for pleasure and doing homework and corresponding increases in television use.[61]

Lessons from the First Wave

The ongoing debate about the design and desirability of standards-based reform hinges critically on how such policies may influence a variety of outcomes among students with different backgrounds. In this chapter, I provided new evidence on those issues by examining the effects of the earlier state-level standards on several outcome and process measures. These results demonstrated that the first wave of student-level standards appears to have had many of both the positive and negative effects suggested by commentators on both sides of these issues. For example, these reforms led to reductions in educational attainment that were particularly large for black students. Furthermore, minimum competency testing led to some apparent reductions in curricular effort while higher course graduation requirements had negative effects on the amounts of time students spend watching television, doing homework, and

reading for pleasure. However, these reforms also increased subsequent employment probabilities. And higher CGRs were partly responsible for the substantial academic upgrading of high school curricula that occurred over this period. In light of this mixed evidence, what can these prior state-level experiences contribute to the current discussions about standards?

A productive, though modest, initial step may be to consider what these results would suggest, to a proponent of standards-based reform, about how those standards should be designed. In particular, the first-wave reforms provide an interesting basis for comparison because they included both a test-based standard (MCT) and a process standard (CGRs). The results presented here suggest that advocates of standards-based reform may prefer the ultimate effects of process standards to those of a test-based standard. More specifically, minimum competency testing had relatively few of the desired effects on educational attainment and early labor market experiences (tables 10-2 and 10-3). The results are consistent with the widely held perception that test-based standards were often weak because of political pressures and the relatively easy and veiled manner in which they could be subsequently lowered. In contrast, newly introduced course graduation requirements created more binding, new standards for students and they were also largely immune to subsequent political redesign.

The evidence from student-level transcripts provides additional support for the relative attractiveness of process standards. More specifically, the results in table 10-4 indicate that CGRs contributed directly to the academic upgrading of the high school curriculum over this period. In contrast, this evidence also suggests that the introduction of MCT lowered their curricular effort, particularly in the sciences and mathematics. The one caveat to the comparative attractiveness of CGRs is that their benefits may be attenuated by changes in teacher expectations (for example, how much homework is assigned) and changes in how students allocate their time. Furthermore, whether these comparative, first-wave results have much external validity for ongoing efforts to develop test-based accountability is clearly open to conjecture. But, at a minimum, the early state-level experiences with minimum competency testing provide an important, cautionary tale.

The implications of the results presented here for the broader debate over whether standards are a desirable type of education reform must be based on more subjectively normative grounds. For example, an advocate of a Rawlsian social welfare function would almost certainly look with favor on such reforms because they increase the employment probability of those who would have dropped out of school anyway. However, others with a more utilitarian perspective may be less willing to accept small employment gains for

many students at the cost of significant welfare losses among those encouraged by new standards to drop out of high school. Those welfare losses encourage advocates of standards-based reform to recommend the simultaneous adoption of targeted efforts to assist those who may be newly at risk of dropping out. Similarly, critics of standards also suggest that, if standards are to be implemented, they should be accompanied by increased capacity building in the form of higher teacher salaries, teacher training, and local control of schools.

However, such approaches to attenuating the difficult trade-offs implied by higher standards may provide a deceptively facile solution. In particular, a fundamental motivation for instituting standards in the first place has been the controversial claim that educational inputs cannot be targeted in ways that systematically promote student achievement. So, recommendations to help somehow the students harmed by standards bring a return to the notoriously difficult research questions about which programs or expenditures might be effective.

The experiences within some of the states that adopted first-wave reforms suggest no one should be too sanguine about their ability (or willingness) to craft solutions that soften these difficult trade-offs. For example, consider the first-wave reforms that were introduced in the two states with the largest public school enrollments. The state of California instituted a new course graduation requirement (first effective for the graduating class of 1987) as part of the Hughes-Hart Educational Reform Act of 1983 (Senate Bill 813). This legislation was a comprehensive school reform package that combined a new state CGR with $800 million in new funds targeted at more than eighty other initiatives including higher starting salaries for teachers and a teacher mentoring program.[62] Similarly, the state of Texas introduced minimum competency testing in 1984 (first effective for the graduating class of 1987) as one component of an extensive package of school reforms (House Bill 72). These reforms included a variety of other complementary initiatives such as increased starting salaries for teachers, a teaching career ladder, management training for principals and superintendents, and a "no pass, no play" restriction on extracurricular activities.[63] Some districts in Texas also responded to MCT by developing summer school initiatives targeted at those at risk for dropping out because of the new standards.[64] These examples indicate that many students who faced new state-level graduation standards were also supported by a contemporaneous mix of other financial and regulatory changes. The extra efforts made by the reform states imply that the difficult trade-offs identified in this study are a relatively intractable feature of introducing higher student-level standards. This interpretation suggests that ongoing

public discussions about the desirability of highly centralized standards should explicitly address how those trade-offs might be valued. Furthermore, these discussions should also consider how the diverse set of policy effects presented here might compare with those of alternative proposals such as the second-wave reforms that stress decentralization and local control.

Notes

1. Erik W. Robelon, "An ESEA Primer," *Education Week*, January 9, 2002, pp. 28–29.

2. Lori Meyer and others, "The State of the States," *Education Week*, January 10, 2002, pp. 68–70.

3. Eric Hanushek and Margaret E. Raymond, "The Confusing World of Educational Accountability," *National Tax Journal*, vol. 54, no. 2 (2001), pp. 365–84.

4. Richard J. Murnane and Frank Levy, "Will Standards-Based Reforms Improve the Education of Students of Color?" *National Tax Journal*, vol. 54, no. 2 (2001), pp. 401–15.

5. Alfie Kohn, "Fighting the Tests: A Practical Guide to Rescuing Our Schools," *Phi Delta Kappan*, vol. 82, no. 5 (2001), pp. 349–57; and Murnane and Levy, "Will Standards-Based Reforms Improve the Education of Students of Color?"

6. Kathryn M. Doherty, "Poll: Teachers Support Standards—With Hesitation," *Education Week*, January 11 2001, p. 20.

7. Lynn Olson, "Finding the Right Mix," *Education Week*, January 11, 2001, pp. 12–20; Jay P. Greene, "The Texas School Miracle Is for Real," *City Journal*, vol. 10, no. 3 (2000) (www.city-journal.org/html/10_3_the_texas_school.html [July 9, 2003]); and Peter Schrag, "Too Good to Be True," *American Prospect*, vol. 11, no. 4 (January 3, 2000) (www.prospect.org/print/V11/4/schrag-p.html [July 9, 2003]).

8. Before this, the most extensive public discussions of education reform were occasioned by the *Sputnik* launch in 1957.

9. See, for example, W. James Popham, "The Case for Minimum Competency Testing," *Phi Delta Kappan*, vol. 63, no. 2 (1981), pp. 89–91.

10. These top-down reforms consisted of state-level standards and regulations that influenced teachers (for example, licensing and salaries) and schools (for example, curriculum and length of school day) as well as students (for example, graduation requirements). In contrast, the subsequent, second wave of reform stressed decentralized improvements such as school-based management, teacher professionalism, and school choice. See, for example, Ahmet Saban, "Emerging Themes of National Educational Reform," *International Journal of Educational Reform*, vol. 6, no. 3 (1997), pp. 349–56. The second-wave reforms are sometimes viewed as a response to the failure of the first wave. See, for example, John Chubb and Terry Moe, *Politics, Markets, and America's Schools* (Brookings, 1990).

11. Chris Pipho, "Minimum Competency Testing in 1978: A Look at State Standards," *Phi Delta Kappan*, vol. 59, no. 9 (1978), pp. 585–88.

12. The enactment of this requirement was often delayed for several years to provide students with adequate notice, to allow the states to test and develop their assessments, and to allow schools to adjust curricula in response to court challenges (for example, *Debra* v. *Turlington* in Florida). I identified the existence and effective date of minimum competency testing (MCT) requirements for graduation by drawing on several sources, including Brian A. Jacob, "Getting Tough? The Impact of High School Graduation Exams," *Educational Evaluation and Policy Analysis* (forthcoming); John H. Bishop and Ferran Mane, "The Impacts of Minimum Competency Exam Graduation Requirements on High School Graduation, College Attendance, and Early Labor Market Success," *Labour Economics*, vol. 8, no. 2 (2001), pp. 203–22; and publications from the Education Commission of the States and Lexis-Nexis searches of state newspaper accounts.

13. See James S. Catterall, "Standards and School Dropouts: A National Study of Tests Required for High School Graduation," *American Journal of Education*, vol. 98, no. 1 (1989), pp. 1–34.

14. Robert C. Serow, "Effects of Minimum Competency Testing for Minority Students: A Review of Expectations and Outcomes," *Urban Review*, vol. 16, no. 2 (1984), pp. 67–75.

15. National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Educational Reform* (Government Printing Office, April 1983).

16. The amount of credit for a course is typically defined in terms of Carnegie units, which represent a standardized amount of time spent studying a subject over a full academic year. One unit represents successful completion of a class, which averages fifty minutes per day, five days per week, for 180 school days. See Dean R. Lillard and Philip P. DeCicca. "Higher Standards, More Dropouts? Evidence within and across Time," *Economics of Education Review*, vol. 20, no. 5 (2001), pp. 459–73. *A Nation at Risk* also recommended a one-half Carnegie unit of computer-related courses and two of a foreign language for college-bound students.

17. State-year data on course graduation requirements (CGRs) were drawn from several sources, most of which were published by the Education Commission of the States.

18. Julian R. Betts and Robert M. Costrell, "Incentives and Equity under Standards-Based Reform," in Diane Ravitch, ed., *Brookings Papers on Education Policy 2001* (Brookings, 2001).

19. John H. Bishop, "Nerd Harassment, Incentives, School Priorities, and Learning," in Susan Mayer and Paul Peterson, eds., *Earning and Learning* (Brookings, 1999).

20. Betts and Costrell, "Incentives and Equity under Standards-Based Reform."

21. Serow, "Effects of Minimum Competency Testing for Minority Students"; Murnane and Levy, "Will Standards-Based Reforms Improve the Education of Students of Color?"; and Meredith Phillips and Tiffani Chin, "Comment," in Diane Ravitch, ed., *Brookings Papers on Education Policy 2001* (Brookings, 2001).

22. Claude M. Steele, "A Threat in the Air—How Stereotypes Shape Intellectual Identity and Performance," *American Psychologist*, vol. 52, no. 6 (1997), pp. 613–29;

and "Stereotyping and Its Threats Are Real," *American Psychologist*, vol. 53, no. 6 (1998), pp. 680–81.

23. Murnane and Levy, "Will Standards-Based Reforms Improve the Education of Students of Color?"

24. Philips and Chin, "Comment."

25. See, for example, Norman Frederiksen, *The Influence of Minimum Competency Tests on Teaching and Learning* (Princeton, N.J.: ETS Policy Information Center, March 1994); Lillard and DeCicca, "Higher Standards, More Dropouts?"; Bishop and Mane, "The Impacts of Minimum Competency Exam Graduation Requirements"; John H. Bishop and others, "The Role of End-of-Course Exams and Minimal Competency Exams in Standards-Based Reforms," in Diane Ravitch, ed., *Brookings Papers on Education Policy 2001* (Brookings, 2001), pp. 267–345; Thomas B. Hoffer, "High School Graduation Requirements: Effects on Dropping Out and Student Achievement," *Teachers College Record*, vol. 98, no. 4 (1997), pp. 584–607; Jacobs, "Getting Tough?"; William H. Clune and Paula A. White, "Education Reform in the Trenches: Increased Academic Course Taking in High Schools with Lower Achieving Students in States with Higher Graduation Requirements," *Educational Evaluation and Policy Analysis*, vol. 14, no. 1 (1992), pp. 2–20; and Stanley Legum and others, *The 1994 High School Transcript Study Tabulations: Comparative Data on Credits Earned and Demographics for 1994, 1990, 1987, and 1982 High School Graduates* (Department of Education, 1997).

26. One exception is the difference-in-difference comparisons reported by Frederiksen, *The Influence of Minimum Competency Tests on Teaching and Learning*. Another is Lillard and DeCicca, "Higher Standards, More Dropouts?" (p. 467, model 4), who find that higher CGRs increase the probability of dropping out. However, they do not address the effects of MCT in these fixed effect models.

27. For example, one plausible conjecture is that the states most willing to adopt higher standards are those with an unobserved propensity for higher academic achievement. However, it is similarly plausible to suspect the opposite—that the states with an unobserved propensity for lower student achievement would adopt higher standards more aggressively.

28. Steven Ruggles and others, *Integrated Public Use Microdata Series: Version 2.0* (University of Minnesota, Historical Census Projects, 1997). I excluded respondents born in Nebraska because that state does not use Carnegie units in defining its graduation standards. I identified the year in which each respondent was eighteen by the respondent's age on enumeration day (April 1, 1990).

29. College entrants are those whose highest reported educational attainment was "some college, no degree" or higher. Unfortunately, the Public Use Micro-Data Sample (PUMS) does not distinguish high school graduates from general equivalency diploma (GED) completers. However, to the extent that higher standards generated a reduction in this measure of high school completion, there was not a wholly offsetting increase in GED completion. Furthermore, this sample includes students who attended private schools. However, their inclusion is arguably appropriate because students may switch schools to avoid the consequences of stricter standards.

30. Joshua D. Angrist and William N. Evans. "Schooling and Labor Market Consequences of the 1970 State Abortion Reforms," in Solomon W. Polachek, ed., *Research in Labor Economics* (Stamford, Conn.: JAI Press, 1999), pp. 75–113.

31. Those who report that they are not in the labor force are defined as unemployed to avoid omitting discouraged workers. However, the exclusion of these respondents does not substantively alter the subsequent results.

32. The school attendance variable is meant to control for the fact that those respondents still in school over the last year would have had limited labor market experiences. This specification is similar to those used by Bishop and Mane, in "The Impacts of Minimum Competency Exam Graduation Requirements," on high school graduation, college attendance, and early labor market success. However, I also estimated purer reduced-form versions of this model that excluded the measures of educational attainment and school attendance. The results were similar to those reported here (table 10-3).

33. I report Huber-White heteroscedastic-consistent standard errors, which assume clustered effects specific to each state-of-birth by year-of-birth cell. Also, I found that probit models for the binary dependent variables generated similar results.

34. Some studies represent state CGR policies by the total number of Carnegie units required. However, this measure may more accurately reflect the focus of reform efforts (for example, *A Nation at Risk*), which was to promote course taking in core academic areas.

35. Also, I found that the results were similar in models that matched respondents to the state-year variables by their state of residence five years before the census.

36. Specifically, I estimate, in models with and without state fixed effects, the effects on educational attainment of whether there were any state executions at age eighteen. One virtue of using state executions for this counterfactual is that considerable variation exists over this period both within states and across states. An increasing number of states began executions in 1984.

37. Beverly Duncan, "Dropouts and the Unemployed," *Journal of Political Economy*, vol. 73, no. 2 (1965), pp. 121–34.

38. David Card and Thomas Lemieux, "Dropout and Enrollment Trends in the Postwar Period: What Went Wrong in the 1970s?" in Jonathan Gruber, ed., *An Economic Analysis of Risky Behavior among Youths* (University of Chicago Press, 2001).

39. David Card and Thomas Lemieux find that cohort size is associated with significant increases in pupil-teacher ratios.

40. Card and Lemieux, "Dropout and Enrollment Trends in the Postwar Period"; and Thomas Kane, "College Entry by Blacks since 1970: The Role of College Costs, Family Background, and the Returns to Education," *Journal of Political Economy*, vol. 102, no. 5 (1994), pp. 878–911.

41. The execution of criminals could have an influence on educational attainment because the prosecution of these cases can be a meaningful drain on public resources. However, this reduced-form effect should be small and these models control for real school spending per pupil.

42. The PUMS definition of a high school graduate includes GED completers, so these results may understate the true reform-induced reduction in high school completion.

43. Murnane and Levy, "Will Standards-Based Reforms Improve the Education of Students of Color?"

44. Because MCT did not significantly influence educational attainment among white females, the employment reduction could reflect a lowering of educational effort. The results in the next section provide some suggestive evidence that this may be so by evaluating the effects of these reforms on individual process measures drawn from transcript data and survey questions.

45. The wage regression indicates that the wages of high school graduates are 18 percent higher than those of dropouts. The mean probability of employment among dropouts is 54 percent. The regression model for employment participation suggests that high school graduates are 20 percentage points more likely to be employed. Therefore, the implied premium in expected wages is $0.33w_d$ (that is, 0.74 x 1.18 – 0.54).

46. Given the stated uncertainty about whether the observer will be a dropout or not, the expected wage is based on the probabilities of being a graduate or a dropout (0.86 and 0.14, respectively) and the wage premium for graduates (18 percent). However, these calculations generate similar results if it is assumed that the observer receives only w_d.

47. For black males and females, the policy-induced gains in expected wages are also several times larger than the corresponding losses.

48. See Thomas D. Snyder and Charlene M. Hoffman, *Digest of Education Statistics 2000* (Department of Education, 2001), table 138.

49. Snyder and Hoffman, *Digest of Education Statistics 2000*, table 138.

50. Legum and others, *The 1994 High School Transcript Study Tabulations*.

51. Clune and White, "Education Reform in the Trenches."

52. For example, a pre-reform/post-reform time-series comparison may attribute to higher CGR increases in academic course taking that stem from other time-varying determinants (for example, the growing college wage premium in the 1980s).

53. These reduced-form evaluations do not isolate the structural mechanism for such effects. For example, the evidence that MCT reduced course taking in science and mathematics could reflect a reduction in motivation among more able students or a policy-induced reallocation of school and teacher resources.

54. My National Education Longitudinal Study (NELS) extract does not include all NELS respondents because some of the base-year sample (that is, those in eighth grade in 1988) were not in tenth grade at the time of the 1990 follow-up. However, to be representative of high school sophomores, the extract does include respondents who freshened the sample: tenth graders who were not in eighth grade in 1988.

55. Each course taken by a student was assigned one of the more than twenty-two hundred unique codes associated with the Classification of Secondary School Courses (CSSC). For details of each transcript study, see Legum and others, *The*

1994 High School Transcript Study Tabulations; and Steven J. Ingels and others, *National Education Longitudinal Study of 1988: Second Follow-up Transcript Component Data File User's Manual* (Department of Education, 1995).

56. The sample sizes for the nontranscript outcomes are over nineteen thousand. The lower sample size for course-taking data largely reflects the fact that the schools of some NELS respondents were not surveyed for transcripts. Respondents from Nebraska were also excluded because the state does not use Carnegie units in defining its graduation standards.

57. The restricted use version of High School and Beyond (HS&B) does not directly identify each high school student's state. I identified the states of HS&B schools through a crosswalk of the HS&B data on the 1980, 1981, and 1982 state unemployment rates and published numbers.

58. These results are presented in detail in an earlier conference draft of this chapter, Thomas S. Dee, "Standards and Student Outcomes: Lessons from the 'First Wave' of Education Reform," Harvard University, John F. Kennedy School of Government, June 10, 2002.

59. The reference category is a weak or nonexistent CGR (less than one Carnegie unit required in subject).

60. See the conference draft of this paper for details. These results imply that any subsequent labor market gains associated with higher CGRs could reflect incentive effects as well as sorting effects.

61. See the conference draft of this paper for details. Some ambiguity exists in interpreting these results. It could reflect the direct effect of imposing minimum requirements on student perceptions of the value of learning. However, other issues related to the implementation of new CGRs contribute to these results as well. For example, teachers may assign less homework after increases in state CGRs. Nonetheless, these results suggest that CGR-induced increases in academic course taking were attenuated by undesirable changes in how students allocate their time.

62. Larry Cuban, "School Reform by Remote Control: SB 813 in California," *Phi Delta Kappan,* vol. 66, no. 3 (November 1984), pp. 213–15; and Chris Pipho, "States Move Reform Closer to Reality," *Phi Delta Kappan,* vol. 68, no. 4 (December 1986), pp. K1–K8.

63. Pipho, "States Move Reform Closer to Reality."

64. Edith L. Archer and Judith H. Dresden, "A New Kind of Dropout: The Effect of Minimum Competency Testing on High School Graduation in Texas," *Education and Urban Society,* vol. 19, no. 3 (May 1987), pp. 269–79.