*Review of Educational Research*
*Season XXXX, Vol. XX, No. X, pp. xx–xx*
*DOI: 10.3102/0034654310383147*


# Assessing the Effects of High School Exit Examinations

**Jennifer Jellison Holme, Meredith P. Richards, Jo Beth Jimerson, Rebecca W. Cohen**
*The University of Texas at Austin*

*High school exit exams are affecting a growing majority of high school students. Although exit testing polices were enacted with the goal of improving student achievement as well as postsecondary outcomes, they also have the potential for negative effects. To better understand the effects of exit testing policies, in this article the authors systematically review 46 unique studies that pertain to four domains of expected influence: student achievement, graduation, postsecondary outcomes, and school response. The evidence reviewed indicates that exit tests have produced few of the expected benefits and have been associated with costs for the most disadvantaged students. This review suggests policy modifications that may attenuate some of the negative effects.*

**KEYWORDS:** accountability, exit tests, high-stakes testing.

High school exit exams are affecting a growing majority of U.S. students. As of 2008, 23 states required students to pass an exit exam to receive a diploma, and 3 additional states will institute such requirements by 2012. By 2012, these 26 states are projected to enroll 75% of all public high school students and 84% of all students of color in the United States (Center on Education Policy [CEP], 2008).

Exit exam policies embody a number of underlying goals: They are intended to prod schools to improve the way they educate the lowest achieving students, to increase the effort put forth by students, and to provide certification that students have mastered a given set of skills, thereby increasing the labor market value of a high school diploma. There are concerns, however, that these policies may have unintended adverse consequences, such as increasing drop-out rates for at-risk student populations. This article asks, through a systematic review of the literature, whether exit exam policies produce intended benefits and whether there are any associated costs.

## The Evolution of Exit Testing Policies

Exit testing policies were first adopted in the 1970s in the form of minimum competency exams (MCEs; Resnick, 1980). These tests were adopted during a

time in which global economic trends were creating significant waves in the U.S. economy: The United States was affected by a major global recession with unprecedented "stagflation"—high unemployment and high inflation—as well as increasing offshoring of manufacturing jobs overseas (Sassen, 1990; Supel, 1978). MCEs were perceived as a way to address these economic anxieties by shoring up both educational skills and workforce quality.

These tests spread rapidly in the mid to late 1970s; by 1982, 39 states had passed MCE legislation, and 19 of these states tied the exams to the receipt of a diploma (Linn, Madaus, & Pedulla, 1982; Pipho, 1978, 1983; Resnick, 1980; Winfield, 1990). Because MCEs aimed to strengthen students' basic skills, the tests required students to perform only at a middle school level in a few selected content areas (Dee & Jacob, 2006). These low-level standards came under direct attack in 1983 with the publication of *A Nation at Risk* , which placed much of the blame for the national "educational disarmament" on the focus on minimum skills (National Commission on Excellence in Education, 1983). The ensuing push for higher standards led many states to abandon their MCEs or to increase the skill level required by the tests (Hamilton, 2003).

In the 1990s, exit tests reemerged as a centerpiece of state education policy in the form of more rigorous standards-based exit exams. During this time, the restructuring of the global economy was generating a new set of economic anxieties as high-skill information technology positions were increasingly threatened with offshoring (Sassen, 2000). In response, states adopted new more rigorous exit exams to ensure that students master higher level skills, and, in the words of California's state superintendent Jack O'Connell, graduate "equipped not only to survive in our competitive global economy but to thrive" (California Department of Education, 2006).

By 2015, MCEs will be fully phased out, and students in the 26 states with exit testing policies will be required to pass either a series of end-of-course (EOC) exams, aligned with the content of high school courses, or comprehensive exams, covering skills taught throughout a student's high school career (CEP, 2008; see Figure 1). In some states, these data will be used not only to certify high school credentials but also for the assessment of college readiness (CEP, 2008). Because exit exam policies have been disproportionately concentrated in southern and southwestern states, low-income students and students of color are disproportionately likely to face exit test requirements (Grodsky, Warren, & Kalogrides, 2009; Heubert & Hauser, 1999).

## Understanding the Effects of Exit Exams

The purpose of this review is to evaluate both positive and negative outcomes associated with exit testing policies. To accomplish this goal, we elected to use a "theory of action" framework, through which a policy's underlying assumptions, goals, and targets are used to guide analysis of a policy's outcomes (Malen, Croninger, Muncey, & Redmond-Jones, 2002). To examine a policy under the theory of action frame, it is necessary to identify a policy's articulated premises and unarticulated "internal logics," which is the process by which a policy is presumed to effect change (Malen et al., 2002). Because exit exams seek to influence student outcomes both while students are in high school and beyond, they embody a particularly ambitious array of goals and targets.

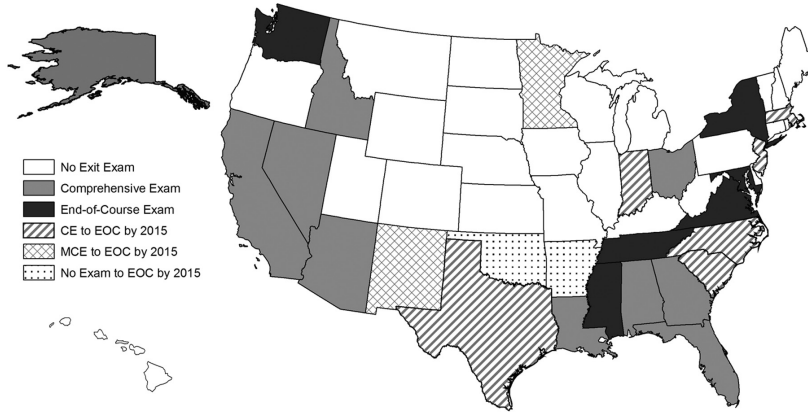


FIGURE 1. *State exit testing requirements by type, current (2009) and projected to 2015. Source.* Center on Education Policy (2009).

*Student achievement.* Exit tests are intended to generate improvements in student achievement by establishing a concrete hurdle that students must clear to graduate. They are also intended to spur achievement by improving the way schools educate their lowest achieving students. A logical question, therefore, emerges: *How have exit tests affected student achievement in high school?*

*Student attainment.* Because exits tests impose an additional graduation requirement, some have expressed concern that exit tests reduce attainment in high school and increase rates of dropout or GED attainment, particularly for at-risk students. These concerns, then, suggest the following question: *How have exit tests affected student attainment in high school?*

*Student postsecondary outcomes.* Exit tests are intended to provide a signal to employers that students with diplomas have mastered a specified set of skills, thereby improving employment and earnings for students with diplomas. By increasing overall achievement, the tests are also expected to lead to increases in college going. These goals prompt a question: *How have exit tests affected student postsecondary outcomes (e.g., college enrollment, employment, and earnings)?*

*School institutional responses.* Exit tests are intended to pressure schools to raise standards to ensure their students pass the exams. However, critics have raised concerns that high-stakes exit tests lead schools to narrow curriculum, particularly for the lowest achieving students. The following question, then, is warranted: *How have schools responded to exit testing requirements?*

## Literature Search

To identify studies for inclusion in this review, we searched major databases (Education Resources Information Center, JSTOR, PsycINFO, Sociological Abstracts, and Social Science Research Network) for articles related to exit exams.

We used the following descriptors in our search: *exit exam(s)*, *exit test(s)*, *exit examination(s)*, and *minimum competency test(s)*. Because minimum competency testing is also used in many professions for employment purposes, in some cases we combined our descriptors with the term *high school* to ensure we included only articles on the subject of high school exit exams. The search began in August 2009 and was updated in January and June of 2010. Together these searches yielded more than 600 results.

### *Inclusion Criteria*

To identify studies for inclusion, we applied the following inclusion criteria, drawn from Guarino, Santibanez, and Daley (2006): relevance, empirical nature, and scholarship quality. With respect to the *relevance* criterion, we included only studies that pertained to the issue of exit testing for a high school diploma and that were relevant to our four central questions. We excluded studies that focused on students with disabilities because this strand of research focused on an extensive, population-specific subset of cognitive and measurement questions that were too broad to include within this review. We also excluded studies that focused on measurement and validity issues related to exit exams.

With respect to *empirical nature*, we included both qualitative and quantitative studies that reported original research rather than the research of others, prior research, or opinion. We also included only studies that met certain criteria for *scholarship quality*, which we based on American Educational Research Association (2006) reporting standards. For inclusion within our analysis, studies must have (a) provided a clear and well-supported logic of inquiry, (b) adequately described the design and sources of evidence, (c) used sources of evidence that were appropriate to the scope of the questions, (d) described the analysis procedures, and (e) described the warrants for conclusions. To ensure the review was as inclusive as possible, we were conservative in our application of these criteria. We therefore included a number of studies that met these reporting criteria yet which drew conclusions that we felt were not warranted by the data or modes of analyses. When applicable, we elaborate our concerns regarding those studies within the text. Furthermore, although we initially restricted our search to studies published in peer-reviewed journals, we found that many rigorous studies with relevant findings had been published as working papers. We therefore determined that publication in a peer-reviewed journal was not a necessary criterion for inclusion our sample.

### *Final Sample and Mode of Analysis*

After application of all inclusion criteria outlined above, 46 unique studies were selected for inclusion in this review, with a total sample of more than 8 million unique students and composite state-level data from all 50 states. We grouped the studies into four separate but overlapping categories based on our research questions. For the student achievement question, we included 16 unique articles, with a total sample of 2,497,536 students and composite state-level data from all 50 states. For the attainment question, we included 23 unique studies, with a sample of 5,265,491 students, composite data from 50 states, and 350 individual districts. For the postsecondary outcome question, we included 9 unique studies with a total sample of 4,688,619 students.[1] Finally, for the school response question, we identified 8 unique studies, with a sample of 271,131 students, 376 teachers, and 49 administrators.

In constructing this systematic review, our goal was to assemble, critically appraise, and synthesize the relevant studies within each of our four categories (Chalmers, Hedges, & Cooper, 2002; Suri & Clarke, 2009). We began our analysis by combing through each of the studies to identify the core questions, sample, and method of analysis. We then clustered studies that addressed similar issues within each of the four categories: for example, within the achievement category we grouped studies that addressed MCEs separately from those that addressed more rigorous exams, and so on. In some instances, significant issues were explored by just a single study; in the interest of presenting a full treatment of the range of issues to consider within a category, we elected to present those studies as a stand-alone cluster for analytic purposes.

We closely analyzed each study's sample and method and considered how each may have affected reported results. To synthesize the findings within cluster, we carefully evaluated how the findings from different studies, years, and data sets related to one another. We then determined whether any reasonable conclusions could be reached about the guiding question for that category.

Within the text, we report the qualitative interpretation of each study's findings, calculated by the authors or in some instances ourselves. In the accompanying tables we report sample sizes, effect sizes (standardized whenever possible), standard errors, and *p* values. Throughout the text, we offer our interpretations for any discrepancies or inconsistencies in the findings.

## Exit Testing and Student Achievement

High school exit exams could be expected to positively influence student achievement through a number interrelated pathways. For example, by pressuring teachers to align curriculum to state content standards, the tests could increase the rigor of the content that students are learning. In addition, the information provided by the test could help teachers identify students who are achieving below the bar set by the state and who need additional support. The threat of diploma denial may also motivate students to work harder to reach the standards set by the state.

There is a possibility, however, that high school exit exams may have little influence on student achievement, particularly if students fail to receive academic supports they need to improve. There is also the potential for exit tests to *negatively* influence student achievement for those students who are unsuccessful on their initial attempt. Instead of rising to meet the challenge, these students may become discouraged and reduce their academic effort.

### *Evidence on MCEs and Student Achievement*

Because MCEs test students in middle-school-level skills, there is reason to expect that these tests would have little impact on the achievement of average or high-performing students, and the evidence has generally supported this assumption (see Table 1). Bishop, Mane, Bishop, and Moriarty (2000) examined this issue using the National Education Longitudinal Study of 1988 (NELS:88) data set, which tracked students enrolled in the 8th grade in 1988 through high school and several years postgraduation. Students participating NELS:88 filled out detailed surveys about background characteristics and were tested in reading, mathematics, social studies, and science during three waves of data collection (when students were enrolled in the 8th, 10th, and 12th grades). Bishop et al. compared test score

**TABLE 1**
*Exit exams and student achievement*

| Author | Data set | Result and effect size |
|---|---|---|
| Minimum competency exams | | |
| Bishop, Mane, Bishop, and Moriarty (2000) | NELS:88<br>$N$ = 10,762–16,153 | No overall effect of MCEs on test score gains for students<br>Significant interaction between MCEs and GPA on test score gains ($\beta = -0.252$, $SE = 0.15$, $p < .05$)<br>—MCEs associated with test score gains for students with C– GPAs ($\beta = 0.42$, $p < .1$)<br>—MCEs not associated with test score gains for students with A or B GPAs |
| Jacob (2001) | NELS:88<br>$N$ = 9,016 | No overall effect of MCEs on 12th grade math or reading scores<br>MCEs associated with smaller gains on reading achievement in MCE states than non-MCE states for students in the bottom decile on reading ($\beta = -1.859$, $SE = 0.665$, $p < .05$) |
| More rigorous exams | | |
| Grodsky, Warren, and Kalogrides (2009) | NAEP LTT 1971–2004<br>13-year-olds:<br>—Math: $n$ = 82,407<br>—Reading: $n$ = 139,578<br>17-year-olds:<br>—Math: $n$ = 70,801<br>—Reading: $n$ = 122,497 | Exit tests (MCE or more difficult exams) not associated with achievement in reading or math at any level of achievement distribution for 13- or 17-year-olds |

*(continued)*

**TABLE 1 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| Reardon, Atteberry, Arshan, and Kurlaender (2009) | 3 cohorts of 10th graders in 4 CA districts (2005–2007)<br>$N = 64{,}226$ | No overall effect of exit exam on student achievement from Grades 10 to 11<br>No effect of exit exam on students in bottom quartile of achievement distribution<br>Blacks, Hispanics, females, and ELLs underperformed on exit test vis-à-vis White students:<br>—Black: $\beta = -0.156$, $SE = 0.025$, $p < .001$<br>—Hispanic: $\beta = -0.108$, $SE = 0.023$, $p < .001$<br>—Female: $\beta = -0.049$, $SE = 0.011$, $p < .001$<br>—ELL: $\beta = -0.082$, $SE = 0.013$, $p < .001$ |
| Impact of failing exit exam | | |
| Reardon, Arshan, Atteberry, and Kurlaender (2008) | 3 cohorts of 10th graders in 4 CA districts (2004–2006)<br>$n = 25{,}299$ (Math)<br>$n = 25{,}596$ (ELA) | "Barely failing" the ELA or math CAHSEE has no effect on subsequent student achievement |
| NAEP 4th and 8th grade | | |
| Bishop, Mane, et al. (2000) | NAEP<br>State-level analysis | States with MCEs had higher NAEP scores than non-MCE states on:<br>4th-grade reading<br>—1992: $b = 3.64$, $SE = 1.43$, $p < .05$<br>—1994: $b = 4.64$, $SE = 2.35$, $p < .05$<br>—1998: $b = 3.35$, $SE = 2.24$, $p < .1$ |

*(continued)*

**TABLE 1 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| | | 4th-grade math<br>—1992: $b = 3.24$, $SE = 1.43$, $p < .05$<br>—1996: $b = 4.18$, $SE = 1.88$, $p < .05$<br>8th-grade math—1992: $b = 2.94$, $SE = 1.53$, $p < .05$<br>—1996: $b = 2.71$, $SE = 1.78$, $p < .1$<br>States with MCE-EOC hybrids had higher NAEP scores than non–exit test states on:<br>4th-grade reading<br>—1992: $b = 4.53$, $SE = 2.37$, $p < .05$<br>—1994: $b = 7.35$, $SE = 3.17$, $p < .05$<br>—1998: $b = 9.9$, $SE = 5.7$, $p < .01$, one-tailed test<br>8th-grade reading<br>—1998: $b = 7.74$, $SE = 2.4$, $p < .01$<br>4th-grade math<br>—1996: $b = 9.7$, $SE = 5.1$, $p < .05$, one-tailed test<br>8th-grade math<br>—1996: $b = 5.56$, $SE = 2.93$, $p < .05$, one-tailed test<br>The authors provide insufficient data to determine standardized measure of effect size; coefficients expressed in terms of point differences – 11.5 points = 1 GLE |
| Amrein and Berliner (2002) | NAEP<br>State-level analysis | Compared to NAEP national average composite scores, states with high-stakes:<br>—Outperform: 4th grade—8 states; 8th grade—5 states<br>—Underperform: 4th grade—3 states; 8th grade—4 states<br>—Perform equally well: 4th grade—2 states; 8th grade—1 state<br>Correlation between state gains and exclusion rates ($R^2 = .0$ to $.49$) |

*(continued)*

**TABLE 1 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| Rosenshine (2003) | NAEP<br>State-level analysis | Compared to non-high-stakes states, high-stakes states had moderate to large advantages in gains in:<br>—4th-grade math: Cohen's $d = 0.35$<br>—8th-grade math: Cohen's $d = 0.79$<br>—4th-grade reading: Cohen's $d = 0.61$ |
| Braun (2004) | NAEP<br>State-level analysis | Compared to low-stakes states, high-stakes states had large advantages in gains in math:<br>—4th-grade: Cohen's $d = 1.08$<br>—8th-grade: Cohen's $d = 0.93$<br>Change in percentage excluded not associated with changes in NAEP scores:<br>—4th-grade: $\beta = 0.1$, $SE = 0.3$, *ns*<br>—8th-grade: $\beta = 0.3$, $SE = 0.3$, *ns* |
| Amrein-Beardsley and Berliner (2003) | NAEP<br>State-level analysis | Accounting for exclusion rates ("unclear" gains or losses), no differences between high-stakes and low-stakes states on 4th-grade reading or math<br>Accounting for exclusion rates, high-stakes states exhibited larger gains than low-stakes states (Cohen's $d = 0.90$, $p < .01$; computed from data) |
| SAT and ACT | | |
| Amrein and Berliner (2002) | ACT/SAT<br>State-level analysis | ACT: Compared to national average, after exit exams were implemented, high-stakes states exhibited losses on ACT 2 × more often than national average (6 states +, 12 states –)<br>SAT: Compared to national average, after exit exams were implemented, high-stakes states exhibited losses on SAT 1.3 × more often than national average (8 states +, 10 states –) |

*(continued)*

**TABLE 1 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| Marchant and Paulson (2005) | NAEP CCD<br>State-level analysis<br>SAT<br>Student-level analysis<br>$N$ = 694,900 | Graduation exams negatively related to state composite SAT score (semipartial = –.099, $p < .05$)<br>Graduation exams negatively related to individual student SAT scores (semipartial = –.039, $p < .001$) |
| International | | |
| Bishop (1998) | TIMSS-95<br>$N$ = 40 countries | Countries with CBEEEs exhibited:<br>—1.3 U.S. GLE advantage in science ($p < .01$)<br>—1.0 U.S. GLE advantage in math ($p < .08$) |
| Woessmann (2003) | TIMSS-95 and TIMSS-Repeat (1999)<br>$N$ = 447,089 students in 77 countries | Countries with exit exams associated with higher scores on TIMSS:<br>Math<br>—TIMSS-95: $\beta = 0.409$, $p < .01$<br>—TIMSS-Repeat: $\beta = 0.470$, $p < .01$<br>Science<br>—TIMSS-95: $\beta = 0.397$, $p < .01$<br>—TIMSS-Repeat: $\beta = 0.359$, $p < .01$ |
| Woessmann (2004) | TIMSS-95 and TIMSS-Repeat (1999)<br>$N$ = 447,089 in 77 countries<br>PISA (2000)<br>$N$ = 96,758–175,227 in 32 countries | CBEEEs associated with 1 GLE higher achievement on PISA and TIMSS math and science (no coefficients reported)<br>Relationship between CBEEEs and achievement stronger for higher ability students (no coefficients reported) |

*(continued)*

**TABLE 1 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| Woessmann, Ludemann, Schutz, and West (2007) | PISA (2003)<br>Math<br>OECD: $n$ = 219,794 in 29 countries<br>Extended: $n$ = 265,878, in 37 countries<br>Science<br>OECD: $n$ = 118,809 in 29 countries<br>Extended: $n$ = 143,528 in 37 countries | CBEEEs associated with higher student achievement on:<br>Math<br>—OECD: $b$ = 13.7 score points ($p < .10$)<br>—Extended: $b$ = 11.1 score points ($p < .10$)<br>Science<br>—OECD: $b$ = 15.7 score points ($p < .05$)<br>—Extended: $b$ = 13.8 score points ($p < .05$) |
| Huang (2009) | TIMSS<br>$N$ = 276,752 in 36 countries | Students in countries with CBEEEs have higher performance on math, but not science<br>—Math: $b$ = 19.12, $SE$ = 8.32, $p < .05$<br>When countries with CBEEEs include Cyprus and Portugal, CBEEEs not associated with higher performance on math or science |
| Jurges, Schneider, Senkbeil, and Carstensen (2009) | TIMSS Germany<br>$N$ = 4,928 | No difference between German states with and without exit exams on gains in mathematical literacy<br>Exit exams associated with higher performance gains on German-specific curriculum-based test for students in nonacademic track ($b$ = 1.95, $SE$ = 0.033, $p < .001$) |

*Note.* MCE = minimum competency exam; NELS:88 = National Education Longitudinal Study of 1988; NAEP = National Assessment of Educational Progress; LTT = Long Term Trend; ELL = English language learner; ELA = English language arts; CAHSEE = California High School Exit Exam; EOC = end of course; GLE = grade level equivalent; TIMSS = Trends in International Mathematics and Science Study; CBEEE = curriculum-based external exit exam; PISA = Programme for International Student Assessment; OECD = Organisation for Economic Co-operation and Development.

outcomes for students in MCE and non-MCE states based on students' initial 8th grade GPAs, theorizing that the tests would have a greater impact on students in the lower end of the achievement distribution. They found that students with C– GPAs in 8th grade showed significantly higher achievement gains on the NELS assessments (24% of grade level equivalent [GLE] between Grades 10 and 12) than students in non-MCE states. No effects were found for students with A and B GPAs. Although Jacob (2001) also found no differences in achievement gains on the NELS assessments between students in MCE and non-MCE states on average, he did find small *losses* for students who scored in the bottom decile of the NELS reading assessment in the 8th grade: these students, he found, gained .50 of a GLE less between Grades 8 and 12 than similar students in non-MCE states.

Although these two studies concur that MCEs have no effect on the achievement of average- to high-performing students, there remains a lack of clarity regarding the impact on low-achieving students. Furthermore, because these studies focus on tests assessing low-level skills, these findings may not generalize to the newer wave of tests with more rigorous standards (Reardon, Atteberry, Arshan, & Kurlaender, 2009).

### *The Impact of Adopting or Shifting to a More Difficult Exit Exam*

Several researchers have examined whether the shift from MCEs to more rigorous standards-based exit exams, which test students on material learned in high school, has produced hoped-for increases in academic achievement (see Table 1). Grodsky et al. (2009) compared student achievement in states with MCEs and "more rigorous" exit testing requirements using individual-level Long Term Trend (LTT) National Assessment of Educational Progress (NAEP) data from 1971 to 2004. Extracting data on 13- and 17-year-olds from the NAEP math and NAEP reading assessments, they pooled observations across years in states with similar types of exit test requirements to compensate for small sample sizes. They categorized state exit tests as "more difficult" if they included any material introduced in the high school curriculum. Grodsky et al. found no association between the adoption of an exit test (either more or less rigorous) and a shift from a less rigorous to more rigorous test on student achievement. They also did not find any differential effects for students at different levels of the achievement distribution (at the 10th, 20th, 80th, and 90th centiles).

Similar conclusions were reached by Reardon et al. (2009), who analyzed data from four large urban districts in California. They examined differences in student achievement before and after the adoption of the California High School Exit Exam (CAHSEE), a test that is taken in the 10th grade and that is considered a "more rigorous" test because it assesses English language arts (ELA) standards through the 10th grade. Reardon et al. used data on three cohorts of 10th graders, two of which were subject to the exit test requirement and one of which was not. The advantage of within-state data was that they provided a relatively better assessment of the impact of test adoption because there was no potential bias because of unobserved state characteristics. Reardon et al. found that the exit test was not associated with any increase or decrease in student achievement between the different cohorts. Likewise, no association was found for the lowest achieving students.

Taken together, these studies concur that the adoption of more rigorous tests is not associated with any achievement gains or losses. Moreover, the evidence

indicates that low-achieving students—those often targeted by these policies—do not experience gains under the more rigorous exams.

### *The Association Between Stakes and Test Performance by Race, Ethnicity, Language, and Gender*

Some have expressed concerns that the stakes attached to exit tests may generate disparities in test performance by race or gender, insofar as students who would otherwise perform well in low-stakes situations perform worse than predicted on a high-stakes graduation test. This phenomenon is known as "stereotype threat," occurring when test performance is negatively affected by "the threat of being viewed through the lens of a negative stereotype, or the fear of doing something that would inadvertently confirm that stereotype" (Steele, 2007, p. 253). Research has found that stereotype threat negatively affects test performance in testing situations that are "high stakes," particularly when tests are perceived to be diagnostic of intellectual ability (Steele, 2007).

Reardon et al. (2009) investigated this phenomenon by comparing student performance on the high-stakes CAHSEE with performance on the low-stakes California Standards Test (CST). (See Table 1). Consistent with the stereotype threat hypothesis, they found low-achieving Blacks, Hispanics, and English learners underperformed on the CAHSEE by 15.6, 10.8, and 8.2 percentage points, respectively, controlling for performance on the low-stakes CST as well as other background variables. Asians performed worse on the ELA portion only (by 10.7 percentage points), whereas females underperformed on the math portion only (by 9.2 percentage points). Reardon et al. concluded that their findings were consistent with the stereotype threat hypothesis, in that the stakes attached to the exit exam appeared to influence student performance consistent with stereotypes about subgroup ability.

### *The Impact of Failing an Exit Test on Student Motivation to Achieve*

To explore the effect of exit tests on student motivation to learn, Reardon, Arshan, Atteberry, and Kurlaender (2008) examined the effect of failing an exit exam on subsequent student achievement (see Table 1). Exit tests could plausibly influence the subsequent achievement of students who fail in two opposite directions: Those students could be motivated to work harder the following year and improve their achievement, or they could become discouraged and exert less effort, thereby reducing their subsequent achievement. Using a regression discontinuity design, Reardon et al. compared the subsequent achievement of students who scored just below the passing cut score to students just above the cut score on the 11th grade low-stakes CST, controlling for a host of background characteristics. Reardon et al. found that barely failing the high-stakes exit test in 10th grade was not associated with any increase or decrease in subsequent student achievement. In interpreting this finding, the authors concluded that failing the test neither caused students on the margin to become more discouraged nor motivated them to work harder to achieve.

### *Downward Pressure? The Impact of Exit Tests on Fourth and Eighth Grade Achievement*

Researchers have also examined the impact of exit testing policies on student achievement through analyses of Grade 4 and 8 NAEP trend data (see Table 1). Although the linkage between high school exit tests and student achievement in

fourth and eighth grade is questionable, there are several means by which exit tests could theoretically influence achievement in the lower grades. For example, exit tests could induce districts to better articulate and realign their K–12 curriculum or to direct more instructional resources to low-achieving students in the lower grades to prevent them from failing at the high school level.

Bishop, Mane, et al. (2000) analyzed the association between MCEs and composite state NAEP scores in Grades 4 and 8 from 1992 to 1998. They found that MCEs were associated with higher NAEP composite reading scores in fourth-grade reading (.29 to .40 of a GLE advantage depending on the year) and in fourth- and eighth-grade math (.24 to .84 of a GLE advantage depending on the year).

Bishop, Mane, et al. (2000) also examined the impact of more rigorous exit tests on NAEP performance by comparing outcomes of students in MCE states and non-MCE states with those of two states with hybrid MCE-EOC testing systems, North Carolina and New York. In these two states, students could choose to take a more rigorous EOC and receive a diploma acknowledging their achievement. Bishop et al. theorized that EOCs should have increased the rigor across the curriculum, resulting in higher levels of achievement (Bishop, Mane, Bishop, & Moriarty, 2001). Consistent with this hypothesis, they found higher performance (.39 to .86 of a GLE) in the EOC states. Yet because these data are specific to two individual states, it is impossible to determine whether these results were attributable to the test or to unobserved state-level characteristics. They themselves raise this issue, concluding that "that the MCE variable may be proxying for other correlated features of the policy environment" (p. 34), such as school report cards or other rewards or sanctions.

Amrein and Berliner (2002) also examined the association between state exit testing requirements and NAEP performance by comparing composite NAEP score trends in exit test states to national averages. Their analysis found no consistent advantages for states with exit test requirements. Critiquing Amrein and Berliner's use of national averages as the point of comparison, Rosenshine (2003) and Braun (2004) reanalyzed the data and found that states with exit tests experienced larger gains when compared to states with no exit tests. Amrein-Beardsley and Berliner (2003) conducted a similar reanalysis and concluded that any gains by exit test states could not be attributed to exit testing policies because of the high rates of students excluded from NAEP testing in those states. Ultimately, however, few reasonable conclusions about exit testing can be drawn from this set of studies because each failed to control for state-level demographic or policy context variables that may be correlated with gains or losses. This is particularly problematic given that exit tests have been disproportionately adopted by southern states, which tend to have much larger concentrations of poverty as well as racial and ethnic minorities. These unaccounted-for state-level variables may have biased their results in significant ways.

### *The Impact of Exit Exams on College Readiness (SAT and ACT)*

Several scholars have examined the association between exit testing and performance on the SAT and ACT (see Table 1). These studies are based on the presumption that differences in SAT and ACT scores between states with and without exit tests reflect learning gains (or losses) as a result of state exit testing policies (Amrein & Berliner, 2002). In one such analysis, Amrein and Berliner (2002)

compared composite SAT and ACT test performance in states with exit tests against national averages. They found that over the long term states with exit tests lost ground compared to the national average on the ACT (posting losses twice as often as the national average) and on the SAT (posting losses 1.3 times as often as the national average), accounting for participation rates. This study, however, contained several significant flaws, having used national averages as the point of comparison while failing to control for other external policy or demographic variables.

Marchant and Paulson (2005) also examined the association between exit tests and SAT performance using both composite state data as well as individual-level SAT data. They found that exit tests were associated with significantly lower SAT scores at both the state level and the individual level. Their analysis, however, did not control for state- or individual-level poverty; because exit tests are usually in states with high levels of poverty, their findings of a negative association in exit testing states are plausibly spurious.

In addition to the study-specific flaws listed above, concerns have been expressed about the use of college entrance tests as a measure of policy impact, as the population taking these assessments tends to be a select sample of college-bound students. These data sets therefore often do not include the lowest achieving students at whom exit exam policies are often targeted (Clark, Rothstein, & Schanzenbach, 2007). Taken as a whole, then, these studies yield few useful conclusions about the relationship between exit testing and student academic achievement.

### *International Evidence on Exit Testing and Student Achievement*

Researchers also examined the association between exit tests and student achievement by comparing performance on international assessments between countries with and without high school exit exams. These studies have specifically examined the effects of curriculum-based external exit exams (CBEEEs), defined by Bishop (1998) as exams that are subject specific, are set to an external standard, apply to nearly all secondary students, and have consequences for students.

The results from these studies have generally demonstrated positive effects of CBEEEs (see Table 1). Bishop (1998) examined composite seventh- and eighth-grade science and math scores for all 40 Trends in International Mathematics and Science Study (TIMSS) countries. He assigned a value of 1 to countries that had exit test systems and 0 to countries without such systems and used fractional weights for countries that had exit tests in some parts of the country, reflecting the percentage of students affected by exit tests in that country. Bishop found that CBEEEs were positively associated with science performance (a 1.3 GLE advantage) but not mathematics. Similarly, Woessmann (2003) examined student performance on the TIMSS and the TIMSS-Repeat (in which different students were tested on the same items) using the same science and math exit exam weights used by Bishop. Woessmann found that student performance in math and science was substantially higher in countries with CBEEEs than countries without such exams, regardless of the students' level of performance or family background (42.7% of a standard deviation in math and 35.9% of a standard deviation in science). In 2004, Woessmann (2004) extended this analysis by adding the Programme for International Student Assessment (PISA) 2000 database (which includes 15-year-olds from additional countries) to his analysis of the TIMSS and TIMSS-Repeat. He found that students in countries with CBEEEs performed roughly one GLE

better on the TIMSS than their counterparts in countries without exit exams. Using the more recent PISA 2003 data set, Woessmann, Ludemann, Schutz, and West (2007) also found significant student performance advantages on PISA mathematics and science for students in countries with CBEEEs, for both Organisation for Economic Co-operation and Development (OECD) and non-OECD countries.

This group of studies, however, is beset by a number of methodological problems. First is the reliance on TIMSS survey data of 13-year-olds in math and science, many of whom have not yet matriculated to high school, to draw conclusions about the impact of high school exit exams. Although these influences are possible, they are at best only indirect. Second, there is a significant possibility that unobserved country-level differences may explain these performance results. These studies, for example, fail to take into account the tracked nature of many secondary systems, which funnel students into very different high schools based on middle school performance (Huang, 2009). As Huang (2009) observes, the impact of an exam will depend on whether all students, or merely a select academic elite, are subject to the requirement. Finally, and perhaps most significantly, there is a lack of agreement among researchers regarding the countries that have an exam system in place. This problem was noted by Huang, who critiqued Bishop's and Woessmann's characterization of the historically low-performing Portugal and Cyprus as having "no exam," arguing that these countries in fact have exit tests (also see Phelps, 2000). Huang illustrated the impact of this change in categorization by replicating the Woessmann study. Utilizing Woessmann's categorization, Huang found a small positive effect of CBEEEs for math (19 points) but not for science. This effect, however, disappeared when Cyprus and Portugal were reclassified as having a CBEEE. Given these methodological problems, therefore, these studies cannot be used to draw any firm conclusions about the relationship between exit tests and student achievement.

One international study, however, does compensate for many of these problems. Jurges, Schneider, Senkbeil, and Carstensen (2009) examined the impact of exit exams on student achievement within Germany. They compared the performance of academic and nonacademic track students in exit-test and non-exit-test provinces on PISA assessment of mathematical literacy and the PISA assessment of knowledge of the German curriculum. They found that average math achievement was comparable in provinces with and without exit tests. However, they found that exit tests were associated with higher scores on the German curriculum test for students in the nonacademic track. They interpret this to mean that teachers in the nonacademic tracks are more likely to "teach the curriculum if the curriculum is tested" (p. 10).

### *Summary and Analysis*

How do exit tests and, in particular, the shift to more rigorous tests affect student achievement? The most trustworthy evidence in this review points to a lack of overall achievement effects, both for MCEs and for the newer wave of "more rigorous" tests. There is also a noted lack of effect for students at the bottom of the achievement distribution who are the very target of these testing policies. There is also compelling evidence that the high-stakes nature of the test actually reduces performance for racial and ethnic minorities and women in some subjects by activating stereotype threat. Investigations relying on the fourth- and eighth-grade

NAEP, the SAT, and the ACT unfortunately lead to no reliable conclusions because of methodological flaws. Similar methodological problems affect the international studies, which are also affected by a lack of agreement on the existence of exit testing policies in particular countries.

## Exit Testing and Dropout or Graduation

There are reasons to expect that exit tests could increase drop-out rates, particularly for at-risk students. One is the potential discouragement effect of such tests: Failure on an exit test could cause academically marginal students who face other barriers (e.g., economic hardship, retention, pregnancy) to give up and leave school. Another is the potential for schools to pressure students to leave: When schools are evaluated based on exit test results, there is the potential for educators to encourage low-achieving students to drop out or to transfer to a GED program.

On the other hand, some argue that exit tests may not be associated with an increase in drop-out rates because the tests may prevent the graduation of only students who would have likely dropped out for other reasons (Greene & Winters, 2004). There is also the possibility that exit testing policies may actually *reduce* the numbers of dropouts by inducing schools to redirect resources to the lowest achieving students and by motivating low-performing students to work harder and improve (Greene & Winters, 2004).

The findings on this issue depend in part on the treatment and definition of dropout used by the researchers, and therefore, we discuss the definition and the implications as appropriate.

### *Evidence on MCEs on Dropout or Graduation*

Because MCEs are set to a relatively low standard, it is reasonable to assume that these tests should prevent a small number of students from graduating. Consistent with this assumption, a number of studies have found no association between MCEs and graduation and drop-out rates (see Table 2). Using NELS:88 data, Jacob (2001) found that MCEs had no significant association with dropping out before the NELS:88 follow-up survey in 1992. Warren and Edwards (2005) also used NELS:88 and found no association between MCEs and the probability of graduating, dropping out, or obtaining a GED. Similar noneffects on dropping out were found by Dee (2003) using the 1990 census Public Use Microdata Sample (PUMS), which is a 5% sample of cities with more than 100,000 population. Dee concluded that MCEs were not associated with an increase or decrease in the probability of self-reports of high school completion for students overall. Although Bishop and Mane (2001b) did not find that MCEs were associated with higher rates of dropout using NELS:88, they did find that students in MCE states were more likely to take more than 4 years to graduate and were more likely to obtain a GED instead of a diploma.

Although there appears to be no association between MCEs and drop-out rates overall, several studies have found that MCEs *are* associated with increases in drop-out rates for lower achieving students. Jacob (2001), for example, found that students in the bottom quartile on the eighth-grade NELS assessment were significantly more likely (6.5 percentage points) to drop out in states with MCEs. Similarly, Bishop and Mane (2001b) found that students with below average GPAs (C–) in MCE states were 7% less likely to receive a diploma *or* a GED, 13 percentage points more likely to take more than 4 years to graduate, and more likely to

**TABLE 2**

*Exit exams and attainment: The impact of exit tests on dropout, graduation, graduation delay, and GED*

| Author | Data set | Result and effect size |
|---|---|---|
| Minimum competency exams | | |
| Jacob (2001) | NELS:88<br>$N = 12{,}171$ | No overall effect of MCEs on probability of dropping out ($\beta = 0.007$, $SE = 0.007$, *ns*)<br>MCEs associated with increased probability of dropping out for bottom quintile students ($\beta = 0.065$, $SE = 0.028$, $p < .05$; 25% more likely to drop out) |
| Warren and Edwards (2005) | NELS:88<br>$N = 13{,}362$ | No overall effect of MCEs on probability of dropping out or obtaining a GED<br>No interaction between MCEs and achievement or SES on dropout or GED |
| Dee (2003) | 1990 PUMS<br>$N = 1{,}348{,}766$ | No overall effect of MCEs on dropout<br>Negative effect of MCEs on dropout for Black males ($\beta = -0.126$, $SE = 0.0055$ [heteroscedastic consistent], $p < .05$) |
| Bishop and Mane (2001b) | NELS:88<br>$N = 16{,}157$ | Students in MCE states were more likely to<br>—Take more than 4 years to get a diploma ($\beta = 0.489$, $SE = 0.145$, $p < .01$)<br>—Obtain a GED ($\beta = 0.440$, $SE = 0.171$, $p < .01$)<br>Students with below average GPAs in 8th grade in MCE states are more likely to<br>—Have no diploma or GED ($\beta = 0.34$, $SE = 0.026$, $p < .05$)<br>—Take more than 4 years to get a diploma ($\beta = 0.743$, $SE = 0.000$, $p < .01$)<br>—Obtain a GED ($\beta = 0.425$, $SE = 0.022$, $p < .05$) |

*(continued)*

**TABLE 2 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| More rigorous exams | | |
| Dee and Jacob (2006) | PUMS (2000)<br>$N$ = 2,925,005 | Both more and less difficult exams negatively associated with likelihood of completing high school<br>—MCEs: $b$ = –0.005 (percentage points), $SE$ = 0.001, $p$ < .05<br>—More difficult exams: $b$ = –0.007, $SE$ = 0.003, $p$ < .05<br>Negative association between exit exams and completion stronger for Black students<br>—Black males:<br>—MCEs: $b$ = –0.013, $SE$ = 0.005, $p$ < .05<br>—More difficult exams: $b$ = –0.018, $SE$ = 0.008, $p$ < .05<br>—Black females: More difficult exams: $b$ = –0.021, $SE$ = 0.008, $p$ < .05 |
| Warren, Jenkins, and Kulick (2006) | Census CPS: Dropout<br>NCES CCD: Completion<br>American Council on Education: GED<br>$N$ = 1,428 "state years" (51 states × 20 graduating classes) | No significant association between exit exams and dropout rates (CPS)<br>Exit exams associated with lower school completion rates ($b$ = – 0.77 [percentage points], $SE$ = 0.348, $p$ < .05)<br>—MCEs not associated with lower school completion rates ($b$ = –0.09, $SE$ = 0.391, *ns*)<br>—More difficult exams associated with lower school completion rates ($b$ = –2.07, $SE$ = 0.483, $p$ < .01)<br>Exit exams associated with higher rates of GED test taking ($b$ = 0.12, $SE$ = 0.046, $p$ < .01)<br>—MCEs not associated with higher rates of GED test taking ($b$ = 0.09, $SE$ = 0.055, *ns*)<br>—More difficult exams associated with higher rates of GED test taking ($b$ = 0.18, $SE$ = 0.061, $p$ < .01)<br>Association between exit exam policies and completion related to state:<br>—Poverty rate ($b$ = –0.17, $SE$ = 0.060, $p$ < .01)<br>—% Hispanic ($b$ = –0.13, $SE$ = 0.030, $p$ < .01) |

*(continued)*

**TABLE 2 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| Bishop, Mane, Bishop, and Moriarty (2000) | NELS<br>$N$ = 10,762–16,153 | Compared to similar students in non-MCE states, students in New York were<br>—More likely to drop out ($b$ = 0.472, $SE$ = 0.206, $p < .01$)<br>—Twice as likely to get a GED ($b$ = 0.84, $SE$ = 0.29 $p < .01$)<br>—Twice as likely to take more than 4 years to graduate ($b$ = 0.94, $SE$ = 0.23, $p < .01$)<br>Below average achievement students in NY more likely to<br>—Drop out ($b$ = 0.472, $SE$ not reported [n/r], $p < .05$)<br>—Late diploma ($b$ = 1.05, $SE$ n/r, $p < .01$)<br>—Take more than 4 years to get a diploma ($b$ = 0.80, $SE$ n/r, $p < .01$) |
| Bishop, Moriarty, and Mane (2000) | NCES CCD<br>State-level analysis | No significant difference between NY and other states in terms of graduation rates |
| Reardon, Atteberry, Arshan, and Kurlaender (2009) | $N$ = 64,226<br>3 cohorts in 4 CA districts (2003–2005) | CAHSEE associated with lower rates of graduation for all quartiles<br>—Quartile 1: $b$ = –0.088, $SE$ = 0.025, $p < .001$<br>—Quartile 2: $b$ = –0.094, $SE$ = 0.014, $p < .001$<br>—Quartile 3: $b$ = –0.092, $SE$ = 0.020, $p < .01$<br>—Quartile 4: $b$ = –0.153, $SE$ = 0.069, $p < .05$<br>Negative association between CAHSEE and graduation stronger for racial minorities in bottom quartile:<br>—Black: $b$ = –0.191, $SE$ = 0.036, $p < .001$<br>—Hispanic: $b$ = –0.146, $SE$ = 0.021, $p < .001$<br>—Asian: $b$ = –0.173, $SE$ = 0.029, $p < .001$<br>Change in graduation rate for students who fail the initial administration of the test in 10th grade:<br>—Quartile 1: $b$ = –0.201, $p < .001$<br>—Quartile 2: $b$ = –0.082, $p < .001$ |

*(continued)*

**TABLE 2 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| Warren and Jenkins (2005) | Census CPS (1968–2000)<br>Texas and Florida state test data (MCE and CE)<br>$N = 15{,}436$ | No association between moving from a MCT to a more rigorous exam and dropout rate (regardless of GED classification)<br>No evidence that racial/ethnic or socioeconomic inequalities were exacerbated by moving from a MCE to a more rigorous exam. |
| Dee and Jacob (2006) | NCES CCD<br>$N = 10{,}502$ dropout rates<br>350 districts<br>9 academic years | Minnesota BST associated with lower dropout rates in 10th and 11th grades but higher dropout rates in 12th grades:<br>—10: $b = -0.36$ (percentage points), $SE = 0.151$, $p < .05$<br>—11: $b = -0.30$, $SE = 0.175$, $p < .05$<br>—12: $b = 0.33$, $SE = 0.166$, $p < .05$<br>Higher increases in 12th grade dropout for:<br>—High-poverty, high-minority schools: $b = 1.29$, $SE = 0.313$, $p < .05$<br>—Urban schools: $b = 3.13$, $SE = 1.07$, $p < .05$<br>—Rural schools: $b = 0.357$, $SE = 0.212$, $p < .10$ |
| Greene and Winters (2004) | NCES CCD<br>$N = 500$ state years<br>50 states × 10 years | Exit exams not associated with graduation rates using two different graduation definitions<br>Effect of exit exams did not change over time (i.e., shifting from MCE to more difficult exam) |
| Marchant and Paulson (2005) | NCES CCD<br>State-level analysis | States with graduation exams had lower graduation rates (semipartial = –.207, $p < .05$) |
| Haney (2000) | Texas state analysis | Implementation of exit test associated with 50% increase in gap in progression ratios between White and non-Whites<br>—Pre-TAAS: Gap = 14.6 percentage points<br>—Post-TAAS: Gap = 21.5 percentage points |

*(continued)*

**TABLE 2 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| Toenjes and Dworkin (2002) | Texas state analysis | Reanalysis of Haney's data using adjusted progression ratio suggests that increase in pass rates on TAAS were *not* the result of increased dropouts or exemptions |
| Carnoy, Loeb, and Smith (2001) | $N$ = 1,134 high schools in Texas | No association between implementation of TAAS and increases in retention or dropouts (began before implementation)<br>Gap between Blacks and Whites began to narrow after implementation of TAAS |
| Impact of failing exit exam | | |
| Martorell (2004) | $N$ = 505,291 (total), $n$ = 20,711 ("last chance") | Failing last chance test in 12th grade associated with<br>—Reduced likelihood of graduating: $b$ = –0.444, $SE$ = 0.012, $p$ < .05<br>—Increased likelihood of taking GED: $b$ = 0.074, $SE$ = 0.004, $p$ < .05<br>—Increased likelihood of receiving GED: $b$ = 0.044, $SE$ = 0.003, $p$ < .05<br>—Increased likelihood of having no high school credential: $b$ = 0.410, $SE$ = 0.012, $p$ < .05<br>Effect of nongraduation because of failing the test alone:<br>—Overall: $b$ = 0.011 to 0.014 (% of students)<br>—Non-White: $b$ = 0.019 to 0.025<br>—Economically disadvantaged: $b$ = 0.021 to 0.026 |
| Reardon, Arshan, Atteberry, and Kurlaender (2008) | $n$ = 25,299 (Math)<br>$n$ = 25,596 (ELA)<br>3 cohorts in 4 CA districts (2004–2006) | No association between failing either ELA or math 10th grade exit test and graduation rates<br>Failing both ELA and math 10th grade exit test associated with reduced graduation rates ($b$ = –0.052 [percentage points], $SE$ = 0.021, $p$ < .05)<br>Failing 1 of 2 last-chance tests in 12th grade associated with reduced probability of graduating on time:<br>—Pass Math: $b$ = –0.418, $SE$ = 0.037, $p$ < .01<br>—Pass ELA: $b$ = –0.472, $SE$ = 0.040, $p$ < .01 |

*(continued)*

**TABLE 2 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| Ou (2009) | New Jersey state analysis<br>$N$ = 299,948 (2002–2006) | Failing on initial attempt associated with increased likelihood of dropping out (Math: $b$ = –0.011, $SE$ = 0.001, $p$ < .01; Reading: $b$ = –0.005, $SE$ = 0.002, $p$ < .01)<br>Failing on retest associated with increased likelihood of dropping out (Math: $b$ = –0.013, $SE$ = 0.001, $p$ < .05; Reading: $b$ = –0.007, $SE$ = 0.003, $p$ < .05)<br>Racial minority and economically disadvantaged barely failers more likely to drop out |
| Papay, Murnane, and Willett (2008) | Massachusetts state analysis<br>$N$ = 83,892<br>10th graders in 2004 | Failing math exit test associated with lower likelihood of on-time graduation for low-income urban students ($b$ = 0.080 [percentage points], $SE$ = 0.033, $p$ < .05)<br>No effects of failing on suburban or wealthy urban students |
| Psychological outcomes | | |
| Richman, Brown, and Clark (1987) | $N$ = 195<br>NC 11th and 12th graders | High-risk students who failed had lower (4.0%) self-esteem scores after testing (32.8 vs. 31.5, $p$ < .05)<br>High-risk students who passed had lower (10.9%) neuroticism scores after testing (57.6 vs. 51.3, $p$ < .05)<br>High-risk students who failed had higher (25.0%) apprehension scores after testing (4.0 vs. 5.0, $p$ < .01)<br>The authors did not report standard deviations, preventing calculation of standardized measures of effect size. |
| Cornell, Krosnick, and Chang (2006) | $N$ = 911 students in MN | 4.4% of students reported dropping out as a result of failing the exam<br>Students reported negative academic, extracurricular, emotional, and peer impact for 8th–10th graders (no control group) |

*(continued)*

**TABLE 2 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| Catterall (1989) | $N = 58$ administrations and $N = 736$ students in 8 schools | Failing graduation test increases student perceptions of likelihood of dropping out (3-point scale):<br>—All students: $b = 0.889$, $SE = 0.27$, $p < .01$<br>—Students aware they took an MCE: $b = 0.913$, $SE = 0.30$, $p < .01$ |
| Jurges, Schneider, Senkbeil, and Carstensen (2009) | TIMSS Germany<br>$N = 4,928$ | Compared to students not subject to an exit exam, students subject to an exit exam had<br>Higher anger scores:<br>—Nonacademic: $d = 0.24$ ($SD$s; $p < .05$); Academic: $d = 0.21$ ($p < .05$)<br>Higher anxiety scores:<br>—Nonacademic: $d = 0.20$ ($p < .001$); Academic: $d = 0.17$ ($p < .10$)<br>Higher despair scores:<br>—Nonacademic: $d = 0.20$ ($p < .001$); Academic: $d = –0.13$ ($p < .05$)<br>Lower enjoyment with academics:<br>—Nonacademic: $d = –0.18$ ($p < .001$); Academic: $d = –0.18$ ($p < .05$) |

*Note*. NELS:88 = National Education Longitudinal Study of 1988; MCE = minimum competency exam; PUMS = Public Use Microdata Sample; CPS = Current Population Survey; NCES = National Center for Education Statistics; CCD = Common Core of Data; BST = Minnesota Basic Skills Test; TAAS = Texas Assessment of Academic Skills; ELA = English language arts.

obtain a GED instead of a diploma. One study, however, found no effects: Warren and Edwards (2005)—also using NELS:88—found no association between achievement (measured via eighth-grade test scores) and the probability of dropping out for students in MCE states.

Several other studies have found additional heterogeneous effects of MCE on dropping out. Although Dee (2003) found no overall effect using 1990 PUMS self-reports of completion, he did find that MCEs were associated with a 1.26 percentage point reduction in the probability of graduation (a self-reported completion measure) for Black males. In addition, the effect of MCEs on dropping out may depend on the type of district a student attends: Using district-level National Center for Education Statistics (NCES) Common Core of Data (CCD) and the single-year "event" drop-out definition, Dee and Jacob (2006) found that the Minnesota's MCE was associated with an increase in 12th grade drop-out rates statewide (of 0.33 percentage points), with much higher drop-out probabilities in high-poverty, high-minority districts (1.29 percentage points), urban districts (3.13 percentage points), and, to a lesser extent, rural districts (0.36 percentage points.)

Taken together, these studies suggest that MCEs are not associated with an increase in the probability of dropout for average students. However, the evidence does suggest that MCEs may induce low-achieving students and Black males to drop out at higher rates. The evidence from Minnesota shows that students attending urban high-poverty districts appear to be more prone to drop out under these policies, which suggests that urban districts may be responding in negative ways to these policies.

### *The Impact of "More Rigorous" Tests on Dropout or Graduation*

Researchers have sought to examine whether the newer wave of "more difficult" tests (testing material at least at the 9th grade level) has produced higher drop-out rates than the MCEs that they replaced (see Table 2). Studies using national data have generally found that more difficult tests are, in fact, associated with a lower completion rates. Dee and Jacob (2006), using individual-level PUMS 2000 data that relies on self-reports of completion, found that both more and less difficult exams were associated with a lower likelihood of completion (4.0% and 5.5% reduction, respectively). These effects were larger for Black males, who were significantly less likely to complete high school in states with both less and more difficult exams (by 5.2% and 7.3%, respectively), and for Black females, who were 11% less likely to complete high school in states with more difficult exams.

Similar results were found by Warren, Jenkins, and Kulick (2006), who used state-level data from 1975 to 2002 to examine the association between more rigorous tests and dropout. They used three data sets and associated drop-out measures: (a) the Current Population Survey (CPS) state-level status drop-out rate (the percentage of 16- to 19-year-olds without a diploma or GED), (b) the NCES CCD state-level completion rate (a 4-year completion rate based on incoming ninth-graders, adjusted for interstate migration), and (c) state-level GED test-taking rates. Using state and year fixed effects models, they found no association between exit exams and the CPS status drop-out rate. They also found no association between MCEs and the NCES CCD completion rate, but they did find more difficult exams were associated with a 2.1 percentage point reduction in NCES CCD

completion rates. They also found that rates of GED test taking were higher in states with more difficult exams. Examining a number of state-level covariates, Warren et al. found that the negative relationship between exit exams and high school completion was strongest in states with higher poverty rates and higher proportions of Hispanic residents.

The impact of New York's more rigorous test on dropout was investigated by Bishop, Mane, et al. (2000), who used the NELS:88 data set to compare drop-out rates in New York to those in states with MCEs. Although their earlier study using NCES state-level aggregate data (Bishop, Moriarty, & Mane, 2000) found no significant differences in terms of graduation rates, this student-level study found that students in New York were significantly more likely to drop out, twice as likely to get a GED, and twice as likely to take more than 4 years to graduate. Bishop et al. found that these effects were larger for students with low GPAs in the eighth grade, who were more than twice as likely to obtain a diploma late and to earn a GED. However, because this study relied on comparative data from just one state, there is a significant possibility that the results suffer from bias because of unobserved state-level effects.

In addition to these national-level studies, several studies utilizing data from individual states have found that more rigorous tests are associated with higher rates of dropping out, especially for lower achieving students. For example, Reardon et al. (2009) compared different cohorts of 10th graders from four large urban districts in California before and after the institution of the state's exit test requirement. They found that the CAHSEE significantly reduced the probability of graduation for students in the lowest achievement quartile (by 11.2 percentage points) and in the second lowest quartile (by 2.7 percentage points.) For students in the bottom quartile, the effects also differed by race: Although White students experienced no declines, the CAHSEE was associated with reduced graduation chances for Black students (a decline of 19 percentage points), Hispanic students (a decline of 14.6 percentage points), and Asian students (a decline of 17.3 percentage points) compared with prior cohorts. For students who failed the first chance CAHSEE test in the 10th grade, the effects were even larger: Failing was associated with a 20.1 percentage point decrease in the probability of graduating for students in the first quartile and a 8.2 percentage point decrease in the probability of graduating for students in the second quartile.

Warren and Jenkins (2005) analyzed CPS data from Florida and Texas, both of which are states that adopted MCEs in the 1980s and then transitioned to more rigorous exams in the 1990s. In contrast to Reardon et al., they found no association between exit tests and dropouts (defined as the proportion of the preceding year's students who left school and failed to attain a credential). They also found no association for any racial/ethnic subgroup, regardless of whether GEDs were classified as dropouts. However, Warren et al. (2006) later acknowledged that the CPS measure used in this article contained such serious flaws (e.g., inadequate state sample sizes and failure account for interstate migration) as to raise questions about the validity of this article's conclusions (also see Greene & Winters, 2004).

Several other studies were conducted on these questions but included inadequate controls, raising significant doubts about their conclusions. For example, the analysis by Greene and Winters (2004), which found no relationship between graduation exams and drop-out rates, included only two control variables

(student–teacher ratios and state per pupil spending) and neglected to include any demographic controls; as a result, these results were likely biased by unobserved state-level confounds. Similarly, an analysis by Marchant and Paulson (2005), which found that exit tests were associated with reduced graduation rates, neglected to account for any state policy variables that could explain outcomes. Similar problems affected the analysis conducted of Texas's exit testing system by Haney (2000), which debunked claims of increased achievement under the Texas accountability system (the "Texas miracle") by analyzing state-level student progression trends; he found that the adoption of the Texas Assessment of Academic Skills (TAAS) in the early 1990s was associated with a 50% increase in the gap in Grade 9 to Grade 12 progression rates between Whites and non-Whites. This study, along with one study disputing Haney's findings (Toenjes & Dworkin, 2002), did not account for other state-level demographic or policy variables that could have explained these trends. Carnoy, Loeb, and Smith (2001) conducted a school-level regression analysis that also disputed Haney's findings, yet this analysis controlled for limited contextual and school-level variables. Ultimately, then, few definitive conclusions can be drawn from this set of studies.

Taken together, the most methodologically rigorous studies are nearly unanimous in their conclusion that, unlike MCEs, "more difficult" exit tests are associated with increased drop-out rates, delays in graduation, and increased rates of GED attainment. These studies also suggest that the more difficult exams have a particularly adverse impact on outcomes for non-Whites and students residing in both high-poverty states and districts.

### *The Effect of Failing an Exit Exam on Dropout or Graduation*

Some researchers have investigated whether test *failure* influences the probability of a student dropping out of school by comparing outcomes for students who barely failed an exit test (falling just below the cut score) to outcomes for those who barely pass using regression discontinuity methodology. The underlying premise for these studies is that students just below or above the cut score on the test are "observationally similar" in terms of academic ability; therefore, these studies aim to discern the causal relationship between failing a high-stakes test—essentially the effect of the cut score—and discouragement and dropout.

Each of these studies has found that students who barely fail have a higher probability of dropping out than students who make it over the high-stakes hurdle (see Table 2). Martorell (2004), for example, examined the impact of failing Texas's TAAS exam for students near the cut score using data on more than 500,000 students. He found that although barely failing the exit exam in 10th or 11th grade had no impact on drop-out rates, failing the last-chance exam in 12th grade was associated with a reduced likelihood of graduating and increased the likelihood of taking the GED, receiving a GED, and having no high school credential. Martorell estimated that approximately 1% of all students in Texas failed to graduate because they could not pass the exit test and that this effect was larger for non-Whites (1.9% to 2.5%) and economically disadvantaged students (2.1% to 2.6%).

Similar results were documented by Reardon et al. (2008) for students subject to the California exit exam (CAHSEE). In their analysis of student data from four urban districts, they found that students who barely failed both portions of the exit test (math and ELA) in 10th grade had a reduced (–5%) probability of

graduation. More substantial reductions in the probability of graduation were found for students who barely failed the last-chance test in 12th grade on one of the exit test subjects: Students who passed math but failed ELA were 42 percentage points less likely to graduate by the fall after their senior year, whereas students who passed the ELA but failed math were 47 percentage points less likely to graduate by that time.

These findings from Texas and California are consistent with Ou's (2009) results from New Jersey. Ou followed four cohorts of New Jersey students, who are required to take the state's High School Proficiency Assessment in 11th grade, through the end of 12th grade. He found that students who barely failed the initial examination were more likely (1%) to drop out than those who barely passed, as were students who barely failed the retest. He found the negative effects were larger for racial and ethnic minorities and economically disadvantaged students.

Papay, Murnane, and Willet (2010) also found similar results in their study of the impact of the Massachusetts exit exam. They found that barely failing the 10th grade math exam reduced the probability of on-time graduation for low-income urban students by 8 percentage points. Although failing the ELA test was not associated with a reduced probability of on-time graduation, it was associated with a 2 percentage point reduction in the likelihood of graduating within 5 years for low-income urban students. No effect of failing was detected for wealthier students or for suburban students.

All of these studies concur that failing an exit exam is associated with an increased probability of dropping out of high school for students who fall just below the cut score, suggesting that the very experience of test failure on these high-stakes tests leads to discouragement and dropout. These studies collectively suggest that the adverse effects of exit exams may be stronger for low-income students, racial and ethnic minorities, and students residing in urban areas.

### *Psychological Effects*

Researchers have also explored the association between exit tests and dropout by studying the tests' psychological effects on students. Most of these studies have found that students who fail exit exams experience adverse psychological consequences (see Table 2). For example, Richman, Brown, and Clark (1987) surveyed 195 North Carolina 11th- and 12th-grade students before and after taking the state's MCE. They found that, for high-risk students, failing the MCE was associated with small reductions in self-esteem scores (a 4% drop) and substantial increases in apprehension (an 25% increase). High-risk students who passed the MCE, by contrast, showed an 11% decrease in neuroticism scores. Similarly, in face-to-face interviews with more than 700 students, Catterall (1989) found that failing an exit test was associated with a significant increase in doubts about the prospect of graduating (a nearly 1-unit increase on a 3-point scale of perceived chances of dropping out). Catterall's data did not, however, permit any inferences as to whether these reported doubts were related to actual dropout or persistence.

Cornell, Krosnick, and Chang (2006) examined how students were affected when they were mistakenly informed that they had failed the Minnesota exit exam. Cornell et al. surveyed 911 of the 7,898 students who were erroneously informed they failed, only to be told they passed by a small margin several months later. Surveying the students about the experience 2 years after this event, they found that a large majority of students reported adverse emotional reactions, including

depression (64%) and embarrassment (64%). Moreover, 4% of those surveyed reported dropping out of school as a result of this mistaken information.

In their study of the effects of external exit exams in Germany, Jurges et al. (2009) also found negative effects of CBEEEs on self-concept and motivation. Using student-level PISA survey data, Jurges et al. compared the attitudes of students in states with and without CBEEEs and found small but statistically significant differences on measures of anger, anxiety, despair, and enjoyment with academics for all students who were subject to a CBEEE, not just for those students who failed. These effects not only were strongest for students in the nonacademic track but were also significant for students in the academic track.

### *Summary and Analysis*

What can be concluded about the impact of exit tests on student dropout and graduation? Research shows that MCEs have little impact on drop-out rates overall, although there is some evidence that MCEs adversely affect the graduation prospects of low-achieving students and Black males. The evidence on the newer wave of more rigorous tests, however, suggests that more recent exit exam policies are consistently associated with increased drop-out rates, particularly for low-achieving students and for students attending urban high-poverty schools. Research also consistently finds that exit tests induce dropout through discouragement effects: Students who barely fail the test are less likely to graduate than their comparable peers who barely pass, and this is especially true for students in high-poverty urban schools. Studies also suggest that tests may be associated with dropout via negative psychological effects on students.

## Exit Testing and Postsecondary Outcomes

Researchers have also investigated the impact of exit testing requirements on postsecondary outcomes. Below we discuss studies focusing on college enrollment and then examine studies focusing on employment or earnings.

### *The Impact of Exit Exams on College Enrollment*

To the degree that exit tests are successful in improving levels of student achievement, they could indirectly lead to increases in the numbers of students matriculating to college. Yet exit tests may depress college enrollment rates by causing some students to fail to receive diplomas that would have otherwise gone on to college had they not been subject to an exit test requirement.

*MCEs and college enrollment.* Because MCEs are set to such a low standard, it may seem unlikely that MCEs would have any impact on college going. However, if MCEs improve outcomes for the lowest achieving students, there may be some positive association. Bishop and Mane (2001a) examined this issue using High School and Beyond (HSB) data, a longitudinal survey of seniors in the class of 1980 who were followed every 2 years through 1986 (see Table 3). They found that, for students who graduated from high school, MCEs were associated with increased rates of college enrollment, particularly for low-achieving and average students.

Yet although high school *graduates* may have better college-going rates in MCE states, the policy impact may not be positive if the tests reduce the overall number of high school graduates (Jacob, 2001). Dee (2003) examined this issue

**TABLE 3**

*Exit exams and postsecondary outcomes: College enrollment, employment, and earnings*

| Author | Data set | Result and effect size |
|---|---|---|
| College enrollment—MCEs | | |
| Bishop and Mane (2001a) | HSB<br>Sample sizes not reported | MCEs associated with increased likelihood of college enrollment, especially for low- and middle-achieving graduates:<br>—Low achieving: $b = 0.038$, $p < .01$<br>—Middle achieving: $b = 0.045$, $p < .01$<br>—High achieving: *ns* |
| Dee (2003) | PUMS (1990)<br>$N = 1{,}348{,}766$ | Overall, MCEs not associated with likelihood of attending college |
| Bishop and Mane (2001b) | NELS:88<br>$N = 12{,}030$–$12{,}037$ | MCEs associated with increased likelihood of college enrollment 1 year after graduation:<br>—Fall 1993: $b = 0.154$, $SE = 0.079$, $p < .05$; 4% increase in probability<br>—Spring 1994: $b = 0.129$, $SE = 0.079$, $p < .1$; 3% increase in probability |
| Bishop, Mane, Bishop, and Moriarty (2001) | NELS:88<br>$N = 12{,}026$–$12{,}033$ | MCEs associated with higher likelihood of college attendance after 2 years postgraduation:<br>—Fall 1993: $b = 0.177$ (percentage points), $SE = 0.086$, $p < .05$<br>—Spring 1994: $b = 0.140$, $SE = 0.086$, $p < .05$<br>Significant interaction between MCEs and GPA on college attendance; MCE has more positive effects on attendance for high-performing students than low-performing students for 2 years postgraduation ($b = 0.149$ to $0.154$, $SE = 0.076$ to $0.080$, $p < .05$) |

*(continued)*

**TABLE 3 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| College enrollment—More rigorous exams | | |
| Dee and Jacob (2006) | PUMS (2000)<br>$N$ = 2,925,005 | Exit tests (more or less difficult) not associated with probability of college enrollment overall or by race, except<br>—Hispanic females subject to MCE more likely to enroll in college ($b$ = 0.083 [percentage points], $SE$ = 0.027, $p < .05$) |
| Warren, Grodsky, and Lee (2008) | PUMS (1980–2000)<br>$N$ = 1.9 million<br>Census CPS-ORG<br>$N$ = 339,000 | Exit exams (MCE or higher competency) not associated with postsecondary schooling<br>No significant heterogeneity by race/ethnicity |
| Bishop et al. (2001) | NELS:88<br>$N$ = 12,026–12,033 | EOCs (NY) not associated with college attendance |
| Martorell (2004) | Texas state analysis<br>$N$ = 37,114 | Failing last-chance 12th grade test associated with reduced likelihood of<br>—Ever attending 2- or 4-year college: $b$ = –0.057 $SE$ = 0.007, $p < .05$<br>—Ever attending a 4-year institution: $b$ = –0.008, $SE$ = 0.003, $p < .05$<br>Failing last-chance 12th grade test associated with reduced number of<br>—2-year college credits: $b$ = –1.225 (credits), $SE$ = 0.472, $p < .05$<br>—4-year college credits: $b$ = –1.104, $SE$ = 0.228, $p < .05$ |
| D'Agostino and Bonner (2009) | University of Arizona freshmen in 2002–2003<br>$N$ = 2,667 | Of students who took AIMS as sophomores in high school and subsequently enrolling in University of Arizona:<br>Passing AIMS Math test increased predicted 1st year GPA by .43 points ($SE$ = 0.4; β = 0.25, $p < .01$)<br>Passing AIMS Writing test increased predicted GPA by .51 points ($SE$ = 0.03; β = 0.26, $p < .01$)<br>Passing AIMS Reading test had no association with GPA |

*(continued)*

**TABLE 3 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| Earnings and employment—MCEs | | |
| Bishop and Mane (2001b) | NELS:88 and 1994 Follow-Up<br>$N$ = 10,054–11,853 | MCEs associated with 6.7% higher annual earnings ($b$ = 349 [$], $SE$ = 162, $p$ < .05)<br>MCEs associated with 5.0% higher monthly earnings ($b$ = 25.5 [$], $SE$ = 14.7, $p$ < .1)<br>No significant association between MCEs and wage rate, suggesting effects on earnings are the effect of working more hours<br>For below average students, MCEs associated with more months worked ($b$ = 0.92, $SE$ = 0.008, $p$ < .01) |
| Dee (2003) | PUMS (1990)<br>$N$ = 1,348,766 | Overall, MCEs associated with decreased likelihood of being employed ($b$ = –0.0053 [percentage points], $SE$ = 0.0031, $p$ < .1)<br>—Negative effect for White females: $b$ = –0.0095, $SE$ = 0.0031, $p$ < .05<br>—Positive effect for Black males: $b$ = 0.0164, $SE$ = 0.0069, $p$ < .05<br>MCEs not associated with wages |
| Bishop and Mane (2001a) | HSB<br>Sample size not reported<br>NELS<br>Sample size not reported | HSB<br>MCEs associated with higher log wage rates for<br>Students with average test scores in 1981: $b$ = 0.20, $p$ < .10<br>Students with low test scores in 1984: $b$ = 0.41, $p$ < .10<br>Females in 1986: $b$ = 0.034, $p$ < .05<br>MCEs associated with greater earnings (in 1992 dollars) |

*(continued)*

**TABLE 3 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| | | Students with low test scores in 1981: $b = 460$[$], $p < .10$<br>Students with average test scores in 1985: $b = 1077$[$], $p < .01$<br>NELS<br>MCEs associated with higher log wage rates for class of 1992 in 1992–1994<br>Students with average test scores: $b = 0.052$, $p < .01$<br>Students with high test scores: $b = 0.063$, $p < .01$<br>Female students: $b = 0.032$, $p < .10$<br>Male students: $b = 0.41$, $p < .01$,<br>MCEs associated with higher earnings (in 1992 dollars)<br>Students with average test scores: $b = 424$, $p < .05$ |
| Bishop, Moriarty, and Mane (2000) | NELS:88 and Follow-Up HSB<br>*N*s not reported | MCE was associated with higher earnings for female HSB graduates in 2 of the 5 follow-up years:<br>—1982: $b = 0.074$ (percentage points), $SE = 0.036$, $p < .05$<br>—1985: $b = 0.071$, $SE = 0.026$, $p < .01$<br>MCE associated with higher overall earnings for female NELS graduates:<br>—Females: $b = 0.06$, $SE = 0.033$, $p < .1$<br>MCE associated with higher hourly wage rates among all NELS graduates:<br>—Females: $b = 0.034$, $SE = 0.017$, $p < .1$<br>—Males: $b = 0.038$, $SE = 0.018$, $p < .05$ |

*(continued)*

**TABLE 3 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| Earnings and employment—More rigorous exams | | |
| Dee and Jacob (2006) | PUMS (2000)<br>$N$ = 2,925,005 | Overall, exit exams (more or less difficult) not associated with likelihood of employment<br>—More and less difficult exit exams associated with higher likelihood of employment for White females (MD and LD: $b$ = 0.009 [percentage points], $SE$ = 0.005, $p$ < .1)<br>—Less difficult exams associated with lower likelihood of employment for Black males ($b$ = –0.009, $SE$ = 0.005, $p$ < .1)<br>—More difficult exams associated with higher likelihood of employment for Hispanic females ($b$ = 0.031, $SE$ = 0.018, $p$ < .10)<br>Overall, exit exams (more or less difficult) not associated with earnings<br>—Exit exams associated with lower earnings for:<br>—White males (LD: $b$ = –0.014, $SE$ = 0.008, $p$ < .1; MD: $b$ = –0.027, $SE$ = 0.016, $p$ < .1)<br>—White females (MD: $b$ = –0.022, $SE$ = 0.011, $p$ < .05)<br>—Hispanic males (MD: $b$ = –0.071, $SE$ = 0.039, $p$ < .1)<br>—More difficult exams associated with higher earnings for Blacks (Males: $b$ = 0.022, $SE$ = 0.011, $p$ < .1; Females: $b$ = 0.028, $SE$ = 0.015, $p$ < .1) |

*(continued)*

**TABLE 3 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| Warren et al. (2008) | PUMS (1980, 1990, 2000)<br>$N$ = 1.9 million<br>Census CPS-ORG<br>$N$ = 339,000 | Exit exams (MCE or higher competency) not associated with wages<br>Exit exams (MCE or higher competency) not associated with likelihood of being unemployed<br>No significant heterogeneity by race/ethnicity |
| Bishop et al. (2001) | NELS:88<br>$N$ = 10,050–11,849 | EOC exam (NY) associated with 9.7% lower monthly earnings ($b$ = –48 [$], $SE$ = 28.9, $p < .05$)<br>For below average students, EOC exam associated with 27.1% lower monthly earnings ($b$ = –134 [$], $p < .01$)<br>Significant interaction between EOC and GPA on monthly earnings, such that MCEs are associated with more positive consequences for above average students than below average students<br>—Average monthly earnings: $b$ = 40.7 ($), $SE$ = 26, $p < .1$<br>—Spring 1994 monthly earnings: $b$ = 69 ($), $SE$ = 28, $p < .01$ |
| Martorell (2004) | Texas state analysis<br>$N$ = 37,114 | Failing last-chance 12th grade exit exam associated with lower earnings:<br>—Immediately after leaving school: $b$ = –0.025, $SE$ = 0.009, $p < .05$<br>—Effect not significant 2–5 years postgraduation |

*Note.* MCE = minimum competency exam; HSB = High School and Beyond; PUMS = Public Use Microdata Sample; NELS:88 = National Education Longitudinal Study of 1988; CPS-ORG = Current Population Survey Outgoing Rotation Group; EOC = end of course; AIMS = Arizona Instrument to Measure Standards; MD = more difficult; LD = less difficult.

using 1990 PUMS self-reports of educational attainment. He found that MCEs were not associated with an increased or decreased likelihood of attending college. Other studies, however, have found more positive effects: Bishop and Mane (2001b) found, using NELS:88 data, that although MCEs were not associated with increased enrollment immediately after high school, MCEs were associated with increased likelihood (4%) of enrollment in college 1 year after high school graduation. The authors hypothesized that the delays in college enrollment were the result of MCE-related delays in high school graduation, as discussed previously. Using the same data set, Bishop et al. (2001) found that these results varied by GPA: They found that eighth-graders who were A and B students in MCE states were more likely to be delayed in enrolling in college for a year, yet after the delay they were more likely to be enrolled in college than students in non-MCE states (3 percentage points for A students, 4 percentage points for B students). There was no comparable association for average and below average students.

How are we to reconcile these disparate findings? We believe the null findings from Dee's (2003) PUMS analysis to be relatively more reliable because the data account for multiple cohorts and, unlike NELS, account for both state and time-varying characteristics. NELS:88, by contrast, includes only one cohort of students; NELS data with respect to college going are particularly prone to the influence of unobserved state-level and cohort-specific variables that may affect college going for that particular cohort, such as differences in the economic climate and higher education policies between states. These state differences may systematically affect college enrollment for that cohort in unexpected ways. Moreover, Dee's findings are more plausible given that MCEs aligned to sixth-, seventh-, or eighth-grade skills would not be logically expected to be associated with increases in college going overall.

*More rigorous tests and college enrollment.* Researchers have investigated whether the newer wave of more rigorous tests have produced increases in college enrollment, which could occur if many students meet the higher standards and graduate "college ready." Dee and Jacob (2006), for example, examined PUMS Census self-reported data and found no significant impact—positive or negative—of either more or less difficult exit exams on the probability of college attendance. They found one exception for a single demographic subgroup: Hispanic females were more likely to attend college if they lived in a state with a "more difficult" exit exam.

Warren, Grodsky, and Lee (2008) also used PUMS data (1980–2000) in conjunction with the census CPS Outgoing Rotation Group (ORG) survey (1984–2002), a monthly survey of the civilian, noninstitutionalized population. Controlling for state- and student-level demographic covariates as well as covariates related to the state and educational policy environment, Warren et al. found no association between exit tests (either MCEs or higher competency exams) and college attendance and no significant heterogeneity by race/ethnicity.

Similar noneffects were found by Bishop et al. (2001), who examined the effects of New York's EOC system on college attendance rates using NELS:88. They find that New York's EOC system was not associated with increased rates of college attendance. As noted previously, these data consist of one cohort from one state, and therefore, it is difficult to know whether these effects can be attributed to the exit test or to other unobserved state-level variables.

Studies concur, therefore, that more rigorous tests are not associated with increased, or decreased, rates of college enrollment overall. However, Martorell's (2004) study suggests that exit tests can depress college enrollment and attainment for the subset of students who fail to receive a diploma because they barely miss the cut score on the last chance exam at the end of 12th grade. These students, he found, had a significantly reduced likelihood of attending a 2- or 4-year college, and those who attended college earned a lower average number of college credits compared with similar students who barely passed. In total, Martorell concludes that students who fail to attend college as a result of failing the last chance test miss out on 1.3 years of college (40 credit hours) compared to students who barely pass and attend college. It should be noted that Martorell found no impact of failing an exit exam on ultimate attainment of a postsecondary degree. As such, he concluded that the exam primarily affected the college prospects of students who would probably not otherwise have completed their degree requirements.

One study investigated the degree to which one state's "more rigorous" exit test provided an accurate signal to students about their *preparedness* for postsecondary schooling. D'Agostino and Bonner (2009) examined whether the cut score set by the Arizona exit test was predictive of first-year college GPA using data from students at the University of Arizona who had taken the state's exit test but who were not required to pass it for a graduation requirement (thus, not all in the sample had passed the test). They found that the writing and math cut scores were accurate predictors of composite college GPAs; however, they found the cut score on the reading test to be set too low to be a useful predictor of college GPA. Therefore, although more rigorous tests may provide a signal of college preparedness, they do not as yet seem to promote greater rates of college going overall.

### *Summary and Analysis*

What, then, can we say about the effects of exit tests on college enrollment? On the basis of the preceding evidence, we conclude that MCEs are not associated with increased levels of college going for students overall. Although the newer wave of more rigorous tests may provide a somewhat more accurate indicator of college preparedness (D'Agostino & Bonner, 2009), the more rigorous exams seem to have no impact (positive or negative) on college attendance for average students. The evidence from Martorell's study suggests that students who fail the last-chance exit exam have significantly depressed college-going rates but not ultimate degree attainment. Overall, then, there appears to be few effects of these policies on students' postsecondary attainment overall.

### *The Impact of Exit Tests on Labor Market Outcomes*

There are a number of reasons to expect that exit tests could positively affect employment and earnings after high school. One potential way exit tests could influence these outcomes is by certifying or "signaling" that graduates have mastered a certain set of skills, giving them an advantage in both employment and earnings over students without such certification. There is also the potential for inequality in employment and earnings to be reduced because employers inclined to discriminate based on race, ethnicity, or gender may be more willing to hire students who have been certified as possessing a given level of skill and more

willing to pay them equal wages. Furthermore, if exit tests prompt schools to improve instruction and students to work harder, graduates in exit test states may actually have higher skills, which would be theoretically rewarded in the workplace.

On the other hand, exit tests may have no impact on employment or earnings if employers pay little attention to the tests or if employers make hiring decisions based on other factors (e.g., personality, work ethic, social networks). In addition, exit tests may negatively affect employment and earnings for students who fail to obtain a high school diploma as a result of the inability to pass the test, given that a diploma is often a minimum requirement for many entry-level jobs.

*MCEs and labor market outcomes.* Because MCEs are basic skills tests, they would be expected to influence employment and earnings of students who obtain entry-level positions after high school. Bishop and Mane (2001b) used NELS:88 data to investigate whether these tests yielded any employment advantages after high school and found that MCEs were associated with higher annual earnings (6.7%). (See Table 3.) They concluded, however, that these earnings effects were the result of more hours worked by low-achieving (C–) students in MCE states rather than higher wages earned.

Dee (2003) investigated this question using PUMS (1990) data and found that MCEs were associated with a reduced probability of employment overall (a decline of 0.53 percentage points) and for White females (a decline of 0.95 percentage points). Yet Dee (2003) also found that MCEs were associated with positive effects on employment for Black males (a 1.64 percentage point increase). No comparable effects were found for MCEs on wages.

Although these two studies indicate that MCEs have few to no effects on employment for students subject to MCEs overall, studies that have restricted their sample to high school graduates have found some positive, although inconsistent, effects. For example, Bishop and Mane (2001a) analyzed the outcomes of on-time school graduates using the HSB (class of 1980) data set and the NELS:88 (class of 1992) data set. They found heterogeneous results for both the HSB and NELS students: some student subgroups (by race, gender, socioeconomic status [SES], and achievement) were found to have wage rate gains and earnings gains some years but not others. These inconsistent findings were echoed in another analysis of HSB class of 1980 graduates by Bishop, Moriarty, et al. (2000), who found MCEs were positively associated with earnings of female graduates in the HSB sample in some years (7% higher in two of the five follow-up years). Their additional NELS sample results were somewhat consistent: Female graduates had greater earnings overall (6%) and higher hourly wage rates (3%). The effect of MCEs on hourly wage rates was also detected for males (4%).

What are we to make of these findings? These data show no consistent, positive effects for students overall; however, they show sporadic heterogeneous gains for students at different levels of the achievement distribution, for different subgroups (by race, gender, ethnicity), and for specific cohorts. Given that the HSB data (class of 1980) and NELS data (class of 1992) are cohort specific, it is difficult to ascertain whether these effects can be attributed to the exit test rather than other cohort specific variables (e.g., economic or demographic conditions specific to those graduating classes). Given these inconsistencies, our reading of the data is

that the best conclusion that can be drawn is that MCEs had, if not a negative effect, no consistent positive effect on employment or earnings.

*More rigorous tests and labor market outcomes.* The newer wave of rigorous tests could plausibly improve employment and earnings for high school graduates as employers become aware that graduates have met an increased standard, making them more willing to hire and provide greater compensation for those graduates who have met the standard. However, there is also the possibility that employers do not pay attention to the exit test or, even if they are aware, base hiring and compensation decisions on other factors, leading to no improvements.

Using PUMS 2000 data, Dee and Jacob (2006) found no overall association between more or less difficult exit tests and employment, after controlling for attainment (see Table 3). They did, however, find differential effects by race/ethnicity and gender. Specifically, more rigorous tests were associated with slight reductions in the employment of Black males (a decline of 1%) and increases for White females (an increase of 1%) and Hispanic females (an increase of 3%). Likewise, Dee and Jacob found no overall association between more or less difficult exit exams and earnings. Again, however, these effects were heterogeneous by race/ethnicity and gender: They found that states with more rigorous exams were associated with lower earnings for White males (–3%), White females (–2%), and Hispanic males (–7%) but higher earnings for Blacks (an increase of 2% for males, 3% for females).

Warren et al. (2008) analyzed the impact of both MCEs and more rigorous standards-based exit tests on employment and earnings using PUMS (1980–2000) and CPS-ORG (1984–2002) data. Using a sample restricted to 20- to 23-year-olds with no secondary schooling, Warren et al., like Dee and Jacob, found no positive association between exit tests and labor force status, earnings, or wage rates on the overall population. Unlike Dee and Jacob, however, Warren et al. found no variation in outcomes by race or ethnicity or by exam difficulty. Warren et al. conclude that the employment returns to high school completion, and the racial/ethnic disparities in those returns, are not affected by state exit exams and that "the meaning employers attach to a high school diploma is unaffected by state HSEE policies" (p. 97).

Bishop et al. (2001) also used NELS data to examine the effect of the more rigorous New York EOC test on student earnings. Bishop et al. found that the New York exit testing system was associated with 10% lower monthly earnings than non-EOC states overall, although they found advantages for above average students in terms of both monthly earnings (5% to 13% higher) and hourly wage rates (4% higher). Given that the analysis compared one cohort in one state to the other 50 states and Washington, D.C., it is again difficult to attribute these results to the exam itself, as there may have been an unobserved confound within the state environment for that particular New York state cohort.

Martorell's (2004) regression-discontinuity analysis revealed that failing Texas's last-chance test at the end of 12th grade was associated with reduced earnings for those students on the margin of passing who were employed shortly after their graduation. Martorell concluded that because these students had completed all high school credits and were therefore similar in most respects to students who

had passed, diplomas did serve as a signal in the labor market. However, he also found that the earnings disparity was reduced over time, suggesting that the value of the diploma decreased as "labor market participants learn about the true abilities of those who barely failed and barely passed the exam" (p. 34).

*Summary and analysis.* What do we know about the effect of exit tests on labor market outcomes? The findings indicate that exit tests, whether more or less rigorous, have no consistent association with employment or earnings overall. Effects by subgroups (race, ethnicity, gender) are inconsistent and in total suggest few if any positive effects. The Martorell (2004) study does illustrate that *diplomas* do serve as a signal in the marketplace, yet it is unclear whether these signals are a result of exit tests or other factors to which employers are attending.

## School Responses to Exit Testing

Researchers have investigated how schools respond when faced with exit testing requirements. Below we synthesize the results of these studies, focusing first on studies that have examined how schools have changed instruction in response to exit testing pressures, then turning to studies that have examined overall organizational responses.

### *Aligned Instruction or Test Prep?*

A central goal of exit testing policies is to encourage schools to align curricula to the standards set by the state. Concerns have been expressed, however, about the potential for these high-stakes tests to push schools to narrow curricula to material on the test (and to the exclusion of nontested material) or to include explicit "test prep" as part of the curriculum.

Researchers have sought to investigate these potential outcomes (see Table 4). Vogler (2006), for example, surveyed biology teachers in Tennessee about their instructional responses to the state's mandatory EOC biology exit test. He found that instruction was influenced by the EOC: 77% of teachers surveyed reported devoting time to preparing students for the exam; of those, 74% said they devoted 2 or more months to test preparation. However, it was not clear from the survey design whether this test preparation meant increasing coverage of the required curriculum versus a narrowing of curriculum covered.

Direct test prep strategies were more clearly documented in Vogler's 2008 study, in which he surveyed teachers using a set of questions directly asking teachers about time spent on test prep activities (e.g., filling in bubbles, test-taking strategies, practice tests). He surveyed U.S. history teachers in the "high-stakes" context of Mississippi, where the history EOC exam was mandatory for graduation, and teachers in comparatively "low-stakes" Tennessee, where the history EOC was only voluntary (Vogler, 2008). Vogler found that Mississippi teachers spent significantly more classroom time preparing students for the exam: 62% reported spending more than 2 months during instruction incorporating test prep activities, compared to just 14% of Tennessee teachers. In addition, Vogler found that Mississippi teachers' practices were more motivated by exam-related concerns, such as the "format of the accountability examination" and "interest in avoiding sanctions." His results suggest that stakes attached to the test were a key driver of instruction.

**TABLE 4**
*Exit exams and school responses*

| Author | Data set | Result and effect size |
|---|---|---|
| Instructional response | | |
| Vogler (2006) | $N = 141$ biology teachers in TN<br>Stratified random sample via geography and past performance<br>60% response rate | 77% of total sample reported spending time on test preparation<br>Of those, 74.3% spent more than 2 months preparing students for test<br>No association between time spent on test and instructional strategies (teacher vs. student centered) |
| Vogler (2008) | $N = 107$ social studies teachers in MS (64% response)<br>$N = 115$ social studies teachers in TN (63% response)<br>Stratified random sample via geography and past performance | More emphasis on test preparation in high-stakes MS vs. low-stakes TN:<br>Majority of teachers reported spending "some time" on test prep: MS: 83.2%; TN: 54.8%<br>Teachers reporting more than 2 month on test prep: MS: 61.9%; TN: 14.1%<br>Teachers in MS placed greater emphasis on format of exam than TN teachers (83.1% MS vs. 29.4% TN; $\chi^2 = 0.54$, $p < .000$)<br>Teachers in MS reported greater interest in "avoiding sanctions at my school" than TN teachers (69.2% MS vs. 37.4% TN; $\chi^2 = 0.31$, $p < .000$) |
| DeBray (2005) | Case study<br>$N = 15$ (5 math; 8 ELA; 2 admin) | Minimal reports by teachers of instructional change<br>Directly observed test preparation by teachers |
| Holme (2008) | $N = 47$ principals in CA<br>Stratified random sample via initial test scores, oversampling low-performing schools | 11% reported reducing focus on literature; of these, 80% were low-performing schools<br>13% reported adopting explicit test prep strategies; of these, 83% were low-performing schools |

*(continued)*

**TABLE 4 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| Support for students | | |
| Sipple, Killeen, and Monk (2004) | $N = 95$ interviews with 133 individuals in 5 districts | Most prevalent strategy was remediation in core content areas<br>In some instances remediation courses supplanted regular courses |
| Author (2008) | $N = 47$ principals in CA | 40% adopted daytime remediation<br>36% adopted before or after school remediation<br>76% said that students were only given support after students failed the test |
| Serow and Davies (1982) | $N = 1,731$ students from one district in NC | Significant interaction between race and effectiveness of reading remediation; reading remediation was less effective for Black students ($F = 4.18$, $p < .05$)<br>No significant interaction between race and effectiveness of math remediation |
| Burris and Welner (2005) | Longitudinal data on tracked vs. detracked cohorts from Rockville Center NY school district | Membership in cohorts subsequent to detracking was a significantly associated with enrollment in advanced courses and with likelihood of earning a Regents diploma (coefficients and *SE*s not reported, $p < .001$) |
| Adverse response | | |
| Sipple et al. (2004) | $N = 95$ interviews with 133 individuals in 5 districts | 4 of 5 (80%) superintendents and some high school principals (no $N$ reported) reported a "strategy" of shifting students on the verge of dropping out to GED programs, allowing them to count as "transfers" instead of dropouts |

*(continued)*

**TABLE 4 (continued)**

| Author | Data set | Result and effect size |
|---|---|---|
| Vasquez Heilig and Darling-Hammond (2008) | $N$ = 270,000 students over a 7-year period<br>Quantitative<br>$N$ = 160 students, teachers, and administrators at 24 schools<br>Qualitative | Increases in numbers of students who disappeared from enrollment records between 9th grade and completion or dropout coding were associated with increases in 10th grade math test scores ($b$ = 0.082 [percentage points], $SE$ = 0.046, $p$ < .1)<br>Increases in number of students retained in 9th grade associated with increases in TAAS-TLI scores<br>—7% increase in retention associated with a 1-point increase on TAAS-TLI math ($b$ = 0.141 [percentage points], $SE$ = 0.060, $p$ < .01)<br>—8% increase in retention associated with a 1-point increase on TAAS-TLI reading ($b$ = 0.133, $SE$ = 0.051, $p$ < .01) |

*Note*. ELA = English language arts; TAAS-TLI = Texas Assessment of Academic Skills-Texas Learning Index .

Further evidence of test prep was found by DeBray (2005), who conducted a case study of a high-poverty high school's response to New York's voluntary Regents EOC exam. DeBray's interviews with math and English teachers revealed that most teachers were highly committed to the policy's goals of increased curricular rigor. However, ethnographic classroom observations revealed that many the teachers used test questions themselves as a regular part of the curriculum.

Evidence of curricular narrowing was also found by Holme (2008), who interviewed a stratified random sample of school principals in California about their responses to the state's exit exam requirement. Although all of the principals reported increased alignment to state standards in response to the test, some (11%) reported that their teachers were significantly reducing the amount of curriculum covered to focus primarily on tested material. Holme also found that 13% of the principals reported that test preparation strategies were being intentionally integrated into the regular curriculum. These strategies were almost exclusively reported by the low-performing schools in the sample.

These studies provide some evidence that schools are, as intended, better aligning their curriculum and instruction to tested material. Yet there is also evidence that schools, particularly schools that are low performing, engage in curricular "overalignment" by integrating test preparation into the curriculum.

### *Support for At-Risk Students*

In addition to studying how schools change curriculum and instruction, researchers have sought to understand whether exit exams prompt schools to change support systems for at-risk students. From this work, it appears that the dominant response has been the addition of remediation courses for at-risk students. Sipple, Killeen, and Monk (2004), for example, interviewed district administrators and principals about their response to the phase in of the New York Regents exams. They found that the majority of districts instituted some sort of remediation programming, either by adding support courses in a particular subject area or by creating remediation classes in lieu of regular coursework. These findings were echoed by DeBray's (2005) study of a high-poverty high school in the same state: She also found that core English and math departments instituted additional remediation courses for at-risk students.

Similar findings emerged from Holme's (2008) interviews with California high school principals: Although nearly all of the 47 principals reported adopting remediation courses in response to the state's exit test requirement, in practice these remediation courses were not accessible for many students. More than one third (36%) of the 47 principals interviewed reported offering remediation courses before or after school, at a time when many at-risk students had work or family obligations, and as a result few eligible students enrolled in the courses. The majority of principals (76%) also reported that remediation was not offered to students until after they had failed the exam.

Evidence from Serow and Davies's (1982) study suggests that instituting remediation programs for students after they fail the test may not be equally effective for all students. Serow and Davies examined student math and reading score changes as a result of participation in remediation through surveys of 1,731 students who failed the North Carolina MCE in one district. They found that although schools offered remediation equally to all students, this remediation was

less effective for Black students than White students in terms of increasing scores, particularly in the area of reading. They concluded that making remediation equally available to all was not sufficient, given that some students needed more intensive support.

Although many schools respond to exit testing by adopting remediation courses, Burris and Welner's (2005) study of a district in New York illustrates a case in which a district used an exit test (the Regents EOC exam) as a launching point for more dramatic reform. In an effort to ensure that all students had access to the rigorous coursework required to pass the Regents exams, the district eliminated the rigid tracking structure in middle school and in high school. Burris and Welner found that students in the detracked middle school classes, as well as low-SES and special education students in the detracked high school classes, showed sharp improvements in rates of earning a Regents diploma (Burris & Welner, 2005). It should be noted, however, that the article does not report any details of the statistical model.

### *Adverse Organizational Responses*

Researchers have also documented instances in which schools have responded strategically to exit testing requirements in ways that negatively affected students (see Table 4). For example, Sipple et al.'s (2004) study in New York found that both superintendents and principals interviewed "shared a 'strategy' of shifting students on the verge of dropping out to GED programs" (p. 159). This practice allowed students to be counted as transfers instead of dropouts, as state drop-out data at that time did not capture GED transfer rates.

Similar findings emerged from Vasquez Heilig and Darling-Hammond's (2008) mixed-methods longitudinal study of a large urban district in Texas. Both qualitative and quantitative data showed that schools responded to the state's exit test requirement (TAAS) either by encouraging students to drop out or by strategically retaining and skipping students over the high-stakes exit test grade (10th grade). These strategies were found to yield better accountability rankings: Controlling for student demographics, teacher capacity, and other measures, they found that schools with high rates of retentions as well as student "disappearances" had higher exit test scores.

### *Summary and Analysis*

What is known about how schools respond to exit testing requirements? The evidence presented in this section shows that many schools are increasing remediation opportunities for at-risk students, although questions remain as to the degree to which remediation is equally effective or equally available to all students. The literature does show that some schools adopt responses that protect the survival of the organization at the expense of student learning, either through a conscious increase in test prep or, worse, through strategic retention and push out.

## Discussion

Based on our review of the research, we can provide some broad answers to the questions we posed.

How do exit tests and, in particular, the shift to more rigorous tests affect student achievement? The evidence reviewed here indicates that neither MCEs nor

more rigorous tests are associated with increases in student achievement overall or for low-achieving students. Reardon et al.'s (2009) study indicates that the stakes attached to exit tests may actually *reduce* achievement for students for non-Whites, English learners, and females by activating stereotype threat.

What can we conclude about the impact of exit tests on dropping out? Research shows that MCEs are not associated with increased drop-out rates overall. However, more rigorous tests are associated with increased drop-out rates, particularly for low-achieving students, racial minorities, and students in high-poverty urban schools. The increases seem to result at least in part from discouragement: The research finds that students who fall just below the cut score on the test are less likely to graduate compared to their peers who just barely pass. These discouragement effects are largest for racial minorities and students in high-poverty urban schools. Studies also suggest that tests also have negative psychological consequences, which may explain some of the drop-out effects.

How do exit tests affect postsecondary outcomes? Findings indicate that neither MCEs nor more rigorous exams have an impact on college attendance or completion overall. There are, however, negative effects for students who barely fail the exit test at the end of their senior year of high school. Exit tests are not associated with either employment or earnings, suggesting that these tests are not providing signals to employers, who may pay little attention to the tests or may base hiring and compensation decisions on other factors.

How have schools responded to exit testing requirements? The evidence shows that although some schools have engaged in significant reform in response to exit tests, many schools have turned to remediation programs as their major response strategy. Questions remain as to the degree to which remediation programs are equally available to or effective for all students. The evidence does show that some schools are prone to adopting responses that protect the survival of the organization at the expense of student learning.

That exit tests have yielded few, if any, benefits is especially concerning given the costs of these policies to students. Reardon et al. (2009) estimated that each year between 3.6% and 4.5% of high school students in California who have fulfilled all other requirements and have not dropped out before the end of 12th grade fail to receive a diploma as a result of failing the state's test. Although these rates of diploma denial may seem small, the absolute numbers are substantial: At these rates, Reardon et al. (2009) estimate that between 18,000 and 22,500 of California's 10th graders per year who would have otherwise received a diploma will fail to receive a diploma as a result of the exit testing requirement.

Proponents might argue that these tests prevented the graduation of students who had not mastered the requisite skills and that these rates of diploma denial are acceptable costs of this policy. Yet to deem these costs acceptable, they must be weighed against the benefits; thus far, these policies seem to have yielded little in terms of overall student learning, employment, college going, and positive school responses.

## Implications

In terms of implications for policy, this review suggests some ways to ameliorate some of the negative effects associated with the high-stakes nature of the exams. For example, this review suggests that exit testing policies should incorporate alternative ways, beyond retesting opportunities, for students to demonstrate

mastery of state standards. Alternative measures such as portfolio assessments may help to address what Reardon et al. (2009) have shown to be stereotype threat-reduced performance by providing another means, rather than a single high-stakes indicator, for demonstrating achievement. Alternative assessments would also ameliorate some of the discouragement-related drop-out effect that was found for students who barely fail the tests (Ou, 2009; Papay et al., 2010; Reardon et al., 2008) by providing students with an alternate way to demonstrate content mastery. A minority of states with exit testing requirements currently such alternatives in place (CEP, 2009).

Yet although these modifications may moderate some of the costs associated with the policy, the evidence presented in this review indicates that these changes alone are unlikely to generate hoped-for gains in terms of school reform, student learning, or employment outcomes. Additional research is needed to explain variations in school response to these policies (e.g., Burris & Welner, 2005; Vasquez Heilig & Darling-Hammond, 2008); to explore student experiences, particularly with respect to discouragement and stereotype threat; and to understand employers' perspectives on the role of exit tests in employment decisions.

This review has found that exit tests are not as comprehensive a policy solution to the challenge of low student achievement as they have been framed by some of their advocates. Ultimately, improving both achievement and long-term outcomes may involve modifying these fairly straightforward policies in a way that acknowledges the complexity of the problem they purport to solve.

## Notes

We would like to thank the editor of *Review of Educational Research*, Gaea Leinhardt, as well as three anonymous reviewers for their helpful feedback on this article. We would also like to thank Robert Crosnoe for providing feedback on an earlier draft of this article. This research received no specific grant from any funding agency in the public, commercial, or nonprofit sectors.

[1]This excludes sample counts from the studies relying on the High School and Beyond data—Bishop and Mane (2001b) and Bishop, Moriarty, and Mane (2000)—because the sample size was not reported.

## References

American Educational Research Association. (2006). Standards for reporting on empirical social science research in AERA publications. *Educational Researcher*, *35*(6), 33–40.

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, *10*(18). Retrieved from http://epaa.asu.edu/epaa/v10n18

Amrein-Beardsley, A. A., & Berliner, D. C. (2003). Re-analysis of NAEP math and reading scores in states with and without high-stakes tests: Response to Rosenshine. *Education Policy Analysis Archives*, *11*(25). Retrieved from http://epaa.asu.edu/epaa/v11n25/

Bishop, J. H. (1998). The effect of curriculum-based external exit exam systems on student achievement. *Journal of Economic Education*, *29*, 171–182.

Bishop, J. H., & Mane, F. (2001a). The impacts of minimum competency exam graduation requirements on college attendance and early labor market success of disadvantaged students. In G. Orfield & M. L. Kornhaber (Eds.), *Raising standards or*

*raising barriers? Inequality and high-stakes testing in public education* (pp. 51–83). New York, NY: Century Foundation Press.

Bishop, J. H., & Mane, F. (2001b). The impacts of minimum competency exam graduation requirements on high school graduation, college attendance and early labor market success. *Labour Economics*, *8*, 203–222.

Bishop, J. H., Mane, F., Bishop, M., & Moriarty, J. (2000). *The role of end-of-course exams and minimum competency exams in standards-based reforms* (CAHRS Working Paper 00-09). Ithaca, NY: Cornell University, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies.

Bishop, J. H., Mane, F., Bishop, M., & Moriarty, J. (2001). The role of end-of-course exams and minimum competency exams in standards-based reforms. In D. Ravitch (Ed.), *Brookings papers on education policy* (pp. 267–345). Washington, DC: Brookings Institution.

Bishop, J. H., Moriarty, J. Y., & Mane, F. (2000). Diplomas for learning, not seat time: The impacts of New York Regents examinations. *Economics of Education Review*, *19*, 333–349.

Braun, H. (2004). Re-considering the impact of high-stakes testing. *Education Policy Analysis Archives*, *12*(1). Retrieved from http://epaa.asu.edu/epaa/v12n1/

Burris, C. C., & Welner, K. G. (2005). Closing the achievement gap by detracking. *Phi Delta Kappan*, *86*, 594–598.

California Department of Education. (2006, May 30). *Schools chief Jack O'Connell visits Central Valley schools to discuss options for students yet to pass exit exam* [Press release]. Retrieved from http://www.cde.ca.gov/nr/ne/yr06/yr06rel60.asp

Carnoy, M., Loeb, S., & Smith, T. (2001). *Do higher state test scores in Texas make for better high school outcomes?* (CPRE Research Report RR-047). Philadelphia, PA: Consortium for Policy Research in Education.

Catterall, J. S. (1989). Standards and school dropouts: A national study of tests required for high school graduation. *American Journal of Education*, *98*(1), 1–34.

Center on Education Policy. (2008, August). *State high school exit exams: A move toward end-of-course exams*. Washington, DC: Author.

Center on Education Policy. (2009, November). *State high school exit exams: Trends in test programs, alternate pathways, and pass rates*. Washington, DC: Author.

Chalmers, I., Hedges, L., & Cooper, H. (2002). A brief history of the research synthesis. *Evaluation and the Health Professions*, *25*(1), 12–37.

Clark, M., Rothstein, J., & Schanzenbach, D. W. (2009). Selection bias in college admissions test scores. *Economics of Education Review*, *28*, 295–307.

Cornell, D. G., Krosnick, J. A., & Chang, L. (2006). Student reactions to being wrongly informed of failing a high-stakes test: The case of the Minnesota Basic Standards Test. *Educational Policy*, *20*, 718–751.

D'Agostino, J. V., & Bonner, S. M. (2009). High school exit exam scores and university performance. *Educational Assessment*, *14*, 25–37.

DeBray, E. (2005). A comprehensive high school and a shift in New York State policy: A study of early implementation. *High School Journal*, *89*(1), 18–45.

Dee, T. S. (2003). The "first wave" of accountability in No Child Left Behind? In P. E. Peterson & M. R. West (Eds.), *No Child Left Behind? The politics and practice of school accountability* (pp. 215–241). Washington, DC: Brookings Institution.

Dee, T. S., & Jacob, B. A. (2006, April). *Do high school exit exams influence educational attainment or labor market performance?* (Working Paper 12199). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w12199

Greene, J. P., & Winters, M. A. (2004). *Pushed out or pulled up? Exit exams and dropout rates in public high schools* (Education Working Paper 5). New York, NY: Center for Civic Innovation at the Manhattan Institute.

Grodsky, E. S., Warren, J. R., & Kalogrides, D. (2009). State high school exit examinations and NAEP long-term trends in reading and mathematics, 1971–2004. *Educational Policy*, *23*, 589–614.

Guarino, C. M., Santibanez, L., & Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research*, *76*, 173–208.

Hamilton, L. (2003). Assessment as a policy tool. *Review of Research in Education*, *27*, 25–68.

Haney, W. (2000). The myth of the Texas miracle in education. *Education Policy Analysis Archives*, *8*(41). Retrieved from http://epaa.asu.edu/ojs/article/view/432

Heubert, J. P., & Hauser, R. M. (Eds.). (1999). *High stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.

Holme, J. J. (2008). High stakes diplomas: Organizational responses of California's high schools to the state's exit examination requirement. *Research in the Sociology of Education*, *14*, 157–188.

Huang, M. (2009). Do public exams raise student performance? A cross-national difference-in-differences analysis. *Journal of Social Sciences and Philosophy*, *21*(1), 1–21.

Jacob, B. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation and Policy Analysis*, *23*, 99–121.

Jurges, H., & Schneider, K., Senkbeil, M., & Carstensen, C. H. (2009). *Assessment drives learning: The effect of central exams on curricular knowledge and mathematical literacy* (CESifo Working Paper 2666). Munich, Germany: Ifo Institute for Economic Research.

Linn, R. L., Madaus, G. F., & Pedulla, J. J. (1982). Minimum competency testing: Cautions on the state of the art. *American Journal of Education*, *91*(1), 1–35.

Malen, B., Croninger, R., Muncey, D., & Redmond-Jones, D. (2002). Reconstituting schools: "Testing" the "theory of action." *Educational Evaluation and Policy Analysis*, *24*, 113–132.

Marchant, G. J., & Paulson, S. E. (2005). The relationship of high school graduation exams to graduation rates and SAT scores. *Education Policy Analysis Archives*, *13*(6). Retrieved from http://epaa.asu.edu/epaa/v13n6/

Martorell, F. (2004). *Do high school graduation exams matter? A regression discontinuity approach* (Unpublished manuscript). Retrieved from http://www.utdallas.edu/research/tsp/pdfpapers/paper38.html

National Commission on Excellence in Education (1983, April.) A nation at risk: The imperative for educational reform. Washington, DC: Author.

Ou, D. (2009). *To leave or not to leave? A regression discontinuity analysis of the impact of failing the high school exit exam* (CEP Discussion Paper 907). London, England: Center for Economic Performance.

Papay, J. P., Murnane, R. J., & Willett, J. B. (2010). The consequences of high school exit examinations for low-performing urban students: Evidence from Massachusetts. *Educational Evaluation and Policy Analysis*, *32*, 5–23.

Phelps, R. P. (2000). Trends in large-scale testing outside the United States. *Educational Measurement: Issues and Practice*, *19*(5), 11–21.

Pipho, C. (1978, May). Minimum competency testing in 1978: A look at state standards. *Phi Delta Kappan*, *59*, 585–588.

Pipho, C. (1983, January 3). *Student minimum competency testing*. Denver, CO: Education Commission of the States.

Reardon, S. F., Arshan, N., Atteberry, A., & Kurlaender, M. (2008, October). *High stakes, no effects: Effects of failing the California High School Exit Exam* (Working Paper 2008-10). Stanford, CA: Stanford University, Institute for Research on Education Policy & Practice.

Reardon, S. F., Atteberry, A., Arshan, N., & Kurlaender, M. (2009, April 21). *Effects of the California High School Exit Exam on student persistence, achievement, and graduation* (Working Paper 2009-12). Stanford, CA: Stanford University, Institute for Research on Education Policy & Practice.

Resnick, D. P. (1980). Minimum competency testing historically considered. *Review of Research in Education*, *8*, 3–29.

Richman, C. L., Brown, K., & Clark, M. (1987). Personality changes as a function of minimum competency test success or failure. *Contemporary Educational Psychology*, *12*, 7–16.

Rosenshine, B. (2003). High-stakes testing: Another analysis. *Education Policy Analysis Archives*, *11*(24). Retrieved from http://epaa.asu.edu/ojs/article/view/25

Sassen, S. (1990). Economic restructuring and the American city. *Annual Review of Sociology*, *16*, 465–490.

Sassen, S. (2000). Cities in a world economy. In S. S. Fainstein & S. Campbell (Eds.), *Readings in urban theory* (2nd ed., pp. 32–38). Walden, MA: Blackwell.

Serow, R. C., & Davies, J. J. (1982). Resources and outcomes of minimum competency testing as measures of equality of educational opportunity. *American Educational Research Journal*, *19*, 529–539.

Sipple, J. W., Killeen, K., & Monk, D. M. (2004). Adoption and adaptation: School district responses to state imposed high school graduation requirements. *Educational Evaluation and Policy Analysis*, *26*, 143–168.

Steele, C. (2007). Stereotype threat and African American student achievement. In D. B. Grusky & S. Szelenyi (Eds.), *The Inequality Reader: Contemporary and Foundational Readings in Race, Class and Gender* (pp. 252–257). Boulder, CO: Westview Press.

Supel, T. M. (1978, Winter). *The U.S. economy in 1977 and 1978* (Minneapolis Federal Reserve quarterly review). Minneapolis, MN: Minneapolis Federal Reserve.

Suri, H., & Clarke, D. (2009). Advancements in research synthesis methods: From a methodologically inclusive perspective. *Review of Educational Research*, *79*, 395–439.

Toenjes, L. A., & Dworkin, A. G. (2002). Are increasing test scores in Texas really a myth, or is Haney's myth a myth? *Educational Policy Analysis Archives*, *10*(17). Retrieved from http://epaa.asu.edu/epaa/v10n17/

Vasquez Heilig, J., & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis*, *30*, 75–110.

Vogler, K. E. (2006). Impact of a high school graduation examination on Tennessee science teachers' instructional practices. *American Secondary Education*, *35*(1), 33–57.

Vogler, K. E. (2008). Comparing the impact of accountability examinations on Mississippi and Tennessee social studies teachers' instructional practices. *Educational Assessment*, *13*, 1–32.

Warren, J. R., & Edwards, M. R. (2005). High school exit examinations and high school completion: Evidence from the early 1990s. *Educational Evaluation and Policy Analysis*, *27*, 53–74.

Warren, J. R., Grodsky, E. S., & Lee, J. C. (2008). State high school exit examinations and post-secondary labor market outcomes. *Sociology of Education*, *81*, 77–107.

Warren, J. R., & Jenkins, K. N. (2005). High school exit examinations and high school dropout in Texas and Florida, 1971–2000. *Sociology of Education*, *78*, 122–143.

Warren, J. R., Jenkins, K. N., & Kulick, R. (2006). High school exit examinations and state-Level completion and GED rates, 1975–2002. *Educational Evaluation and Policy Analysis*, *28*, 131–152.

Winfield, L. F. (1990). School competency testing reforms and student achievement: Exploring a national perspective. *Educational Evaluation and Policy Analysis*, *12*, 157–173.

Woessmann, L. (2003). Central exit exams and student achievement: International evidence. In P. E. Peterson & M. R. West (Eds.), *No Child Left Behind? The politics and practice of school accountability* (pp. 292–323). Washington, DC: Brookings Institution.

Woessmann, L. (2004). *The effect heterogeneity of central exams: Evidence from TIMSS, TIMSS-Repeat and PISA* (CESifo Working Paper Series 1330). Munich, Germany: Ifo Institute for Economic Research.

Woessmann, L., Ludemann, E., Schutz, G., & West, M. R. (2007). *School accountability, autonomy, choice, and the level of student achievement: International evidence from PISA 2003* (OECD Education Working Paper 13). Munich, Germany: Ifo Institute for Economic Research.

## Authors

JENNIFER JELLISON HOLME, PhD, is an assistant professor of educational policy and planning in the Department of Educational Administration at the University of Texas at Austin, 1 University Station D5400, George Sanchez Building, Austin, TX 78712; e-mail: *jholme@mail.utexas.edu*. Her research focuses on the politics and implementation of educational policy, with a particular focus on the relationship among school reform, equity, and diversity in schools. She received her BA in sociology from UCLA, her EdM in administration, planning and social policy from the Harvard Graduate School of Education, and her PhD in educational policy from UCLA.

MEREDITH P. RICHARDS is a doctoral student in educational policy and planning at the University of Texas at Austin; e-mail: *meredithrichards@gmail.com*. Her research interests focus on the regional and metropolitan context of education, the role of schools in social stratification and segregation, and cognitive perspectives in educational policy implementation. She holds a BA with distinction in cognitive science from the University of Virginia and an MA in industrial/organizational psychology from George Mason University.

JO BETH JIMERSON is a doctoral student in educational policy and planning at the University of Texas at Austin; e-mail: *jbjimerson@gmail.com*. Her research interests focus on data use in educational contexts, teacher development, and issues of student achievement and accountability. She received her BA and teaching certification in English from Texas State University–San Marcos and her MEd in school administration from Trinity University.

REBECCA W. COHEN is a doctoral student in educational policy and planning at the University of Texas at Austin; e-mail: *rweilcohen@gmail.com*. Her research interests focus on the impact of the juvenile justice system on entrants' educational outcomes. She received her BA in sociology from NYU and her master's in middle childhood mathematics education from Brooklyn College.