

Accountability in US Education: Applying Lessons from K–12 Experience to Higher Education

David J. Deming and David Figlio

A new push for accountability has become an increasingly important feature of education policy in the United States and throughout the world. Broadly speaking, accountability seeks to hold educational institutions responsible for student outcomes using tools ranging from performance “report cards” to explicit rewards and sanctions.

In the United States, the accountability movement was presaged by the 1983 publication of *A Nation at Risk*, an incendiary report authored by a commission appointed by President Ronald Reagan. The report famously stated that “if an unfriendly foreign power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war.” *A Nation at Risk* called for—among other changes—more rigorous performance measurement, including nationwide standardized testing. Beginning in the late 1980s, some early-adopting US states—along with countries such as Chile and the United Kingdom—began rating and ranking K–12 schools using measures of student performance. The accountability movement in the United States culminated with the passage of the No Child Left Behind (NCLB) Act of 2001, which required all states to test K–12 students regularly in core subjects and to evaluate

■ *David Deming is Professor, Harvard Graduate School of Education, Cambridge, Massachusetts. David Figlio is Orrington Lunt Professor of Education and Social Policy and Director, Institute for Policy Research, both at Northwestern University, Evanston, Illinois. Deming is a Faculty Research Fellow and Figlio is a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are david_deming@gse.harvard.edu and figlio@northwestern.edu.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <http://dx.doi.org/10.1257/jep.30.3.33>

doi=10.1257/jep.30.3.33

schools based on whether their students were making adequate progress toward achievement benchmarks, with the goal of 100 percent proficiency by 2014.

Oddly enough, we can now view No Child Left Behind Act of 2001 as the beginning of a process of gradual retreat from accountability. Perhaps unsurprisingly, it became clear soon after the passage of NCLB that states would fail to attain the lofty goal of 100 percent proficiency. Starting in 2011, the Obama administration began waiving NCLB requirements for states that agreed to adopt certain policies such as linking teacher evaluations to student test scores. However, this step proved politically contentious and may have contributed to an anti-testing backlash in states such as New York, Florida, and Texas that had formerly implemented some of the most ambitious accountability policies. In December 2015, No Child Left Behind was replaced by the Every Student Succeeds Act, which scales back testing requirements and returns more implementation power to the states.

This dynamic of initial enthusiasm for accountability followed by gradual retrenchment has also played out in US higher education. President Obama announced in 2013 an ambitious plan to rate colleges based on access, affordability, and student outcomes. However, after a two-year process of soliciting feedback from colleges and higher education experts, the Obama administration elected to scrap the rating system as well as any explicit linkage between federal funding and performance. In its place, the administration released the College Scorecard, a database and interactive website where information about graduation rates, earnings, and annual costs of postsecondary institutions can be compared in a standardized format.

In both cases, accountability began with a period of surging policy interest combined with technocratic exuberance about measuring and tracking educational outcomes, which was then followed by caution about unintended consequences and a sense that certain efforts may have overreached. The status quo represents an uneasy compromise. Even the harshest critics of accountability would probably concede that it is hard to stop measuring and tracking performance once you have started. Accountability in some form is probably here to stay. But important questions remain concerning the specifics of how an accountability system should be designed, and what such a system can reasonably be expected to accomplish.

Our purpose in this article is to provide a framework for understanding educational accountability at the K–12 and higher education levels. We start with a discussion of the context from which this push for greater accountability emerged and discuss some of the theoretical arguments behind approaches to accountability in education. We then turn to the well-developed empirical literature on accountability in K–12 education and consider what lessons we can learn for the design and impact of college ratings.

Our bottom line is that accountability works, but rarely as well as one would hope, and often not entirely in the ways that were intended. Research on K–12 accountability offers some hope but also a number of cautionary tales. Importantly, the benefits of K–12 accountability seem to be concentrated among the most disadvantaged students in the lowest-performing schools, both perhaps because failure

is easier to diagnose than success and because lower-performing schools face less scrutiny from stakeholders in the absence of government monitoring.

What lessons can we learn for the design of accountability in US higher education? US colleges and universities vary greatly in selectivity, their broad purposes, and whether they are public or private, making the design challenges even more formidable than in the K–12 setting. In this uncertain environment, the conservative approach of the College Scorecard—standardized performance information, but no explicit stakes—is a sensible choice, though one that still could have significant unintended consequences.

However, we will argue that it is possible to do better by targeting regulatory efforts toward lower-performing institutions of higher education where students have less “skin in the game.” This includes many for-profit colleges that depend heavily on federal financial aid for revenue, but it also potentially includes public (and some private) institutions where few students are paying out-of-pocket due to federal and state subsidies. We think “skin in the game” can work as a guiding principle, for two reasons. First, paying customers send a market signal that schools are providing a valuable product—where value is defined by the student herself rather than policymakers. Second, schools rightly deserve increased public scrutiny when they are more heavily subsidized by taxpayer funds. We close with a call for state-level policy experimentation, in the spirit of critical attention to design details and cautious incrementalism.

The Institutions of School Accountability

The push for school accountability arose from well-known concerns about the performance of the US education system, which we briefly summarize here. At the K–12 level, spending per student has risen substantially over time, from \$5,984 per student in 1970 to \$12,008 in 2000 to \$13,142 per student in 2013 (all expressed in 2015 dollars). However, progress in terms of student achievement has been much slower. Between 1971 and 1999, reading scores of nine year-olds on the National Assessment of Educational Progress (NAEP)—a national exam with consistent scoring over time (it is often referred to as “The Nation’s Report Card”)—rose by only 4 points (from 208 to 212). Gains in mathematics, while stronger, were still relatively modest. Moreover, students from other countries often outperform US students on international tests (as discussed in Woessmann’s paper in this symposium). Encouragingly, NAEP scores have risen relatively rapidly between 1999 and 2012, with particularly large gains for younger students and for students of color. An additional piece of good news is that high school graduation rates rose by more than 10 percentage points between 2000 and 2013 after stagnating during the previous three decades (Murnane 2013). There is some evidence—discussed later in the paper—that accountability has had relatively larger impacts at the bottom of the achievement distribution, and thus may have contributed to the narrowing of achievement gaps over this

period. However, other candidate explanations such as changing family environments and increases in early education lead us to stress that this conclusion is speculative.

Trends in US higher education are much less positive. While college attendance rates have risen steadily over the last several decades, high dropout rates have led to only modest growth in bachelor's degree attainment. Only about 60 percent of bachelor's degree-seeking students successfully obtain a degree after six years. As a result, the US four-year college degree attainment rate ranks slightly below the average of high-income OECD countries.

Public higher education in the United States is funded primarily by a combination of student tuition and state legislative appropriations. State subsidies allow public colleges and universities to spend more per student than they charge in tuition prices. Yet declining state support means that students are paying for a larger share of their education. Between the 1999–2000 and 2014–2015 school years, inflation-adjusted state funding per full-time equivalent (FTE) student declined by about 25 percent. Tuition revenue per FTE student increased over this period by a similar amount, leaving total per-student spending in public institutions relatively constant. Prices are increasing, but spending is not—students are just footing a higher share of the bill through out-of-pocket spending and student loans, which now total more than \$1.3 trillion.

Despite these headwinds, college continues to be a worthwhile investment on average. Avery and Turner (2012) estimate that the present discounted lifetime value of a college degree relative to a high school degree—net of tuition—is positive and large, and has actually grown over time despite rising prices and growing student loan debt. In sum, college appears to be an increasingly risky—yet also increasingly necessary—investment.

What is the role of US educational institutions in producing these mixed outcomes? The governance and funding of the US education system, as in most countries, comes primarily from the public sector. US K–12 public schools are managed by elected school boards, while principals, teachers, and other employees are public sector workers. US higher education has a more diverse institutional structure, but the majority of students attend public institutions that receive considerable public support, both directly through tax revenues and indirectly through provision of student loans and other methods. Private colleges and universities receive considerable government funding as well, be it directly through program support or grants or indirectly through subsidization via the deductibility of charitable contributions that subsidize institutional endowments, and students of public and private institutions alike receive federal and state financial aid. Most institutions of higher education, whether public or private, are managed by a combination of a board of trustees, who hire the president and sometimes other top administrators, and the faculty of those institutions.

Given the importance of the public sector in providing and subsidizing education, many issues in educational accountability can be understood through the lens of a classic principal–agent problem. Policymakers, parents, and students wish to

contract with schools to provide education. However, the provision of education requires the system insiders to make an array of decisions and budgetary choices, about hiring, discipline, tenure, curriculum, pedagogy, pay and benefits, grading and exams, and class sizes. It is difficult for interested outsiders to monitor the actions of schools and universities on these and other dimensions. The hope of greater educational accountability is that it will pressure the insiders in schools and universities to alter their production decisions and to improve in some key areas.

When questions arise about improving accountability, an economist's first instinct is often to ask why "the market" cannot provide sufficient accountability among providers. However, as economists have long recognized, education is an industry where the power of consumers to ensure quality by choosing among alternatives is often quite limited: the range of school choices is constrained by political jurisdictions and geography; direct public provision of educational services is widespread, often financed either completely (in the case of K–12 education) or substantially (in the case of higher education) by tax revenues; and possibilities for entry and exit are limited.

In the case of K–12 education, as pointed out by Milton Friedman (1962) more than 50 years ago, the institutional structure does not facilitate competition. Choice among K–12 public schools largely operates indirectly through choice of residential location (for example, Hoxby 2003), although in some places students may have some access to choosing public charter schools or schools in neighboring jurisdictions. Public schools are funded mostly by property taxes.

At first glance, the scope for consumer choice to drive accountability appears much more promising in higher education. Selective colleges and universities compete fiercely for the best students in a nationwide market, and these schools are most often the focus of public discussion. However, the vast majority of US college students attend nonselective and mostly public institutions that are close to home (Hoxby 2009). While public colleges in the United States receive considerable lower levels of state appropriations than once had been the case, many are still heavily subsidized through state legislative appropriations, and most still charge only a fraction of the true per-student cost (Winston 1999). Private colleges are also subsidized through the tax deductibility of charitable contributions, and many of the less-selective private institutions are heavily dependent on federal financial aid subsidies that students bring with them. As a result, the market for higher education—with the possible exception of elite colleges—is probably not very competitive.

The difficulty of fully monitoring actors within the educational system combined with the limited scope for external incentives through consumer choice probably justifies some form of accountability for educational institutions. This can take a variety of forms, from increased information provision and disclosure requirements to more heavy-handed regulation and incentive structures. The hard questions involve figuring out how many institutions are really in need of remediation, as well as the specific design details, including practical and political constraints.

Approaches to Accountability in Education

Educational accountability begins with collecting consistent information on specific outcomes and inputs of interest over time. This information can be used in two broad ways. A first approach, called *report-card accountability*, makes certain information public, but without other explicit stakes. This approach is the norm in many countries. As a consequence of the No Child Left Behind Act of 2001, nearly every US state has developed school report cards with information on test performance and other outcomes by K–12 grade, subject, and student subgroup. The second approach is the use of rewards and sanctions to motivate increased performance—what Hanushek and Raymond (2005) call *consequential accountability*. This means attaching rewards and sanctions to benchmarks, such as the percent of students meeting the proficiency standard on a mathematics test, or the rate of return on investment in a college degree.

The most controversial elements of the No Child Left Behind Act of 2001 were a set of escalating sanctions for repeated failure to meet achievement benchmarks. In the first year a school failed to make “adequate yearly progress,” it was required to develop a school improvement plan. Repeated failure led to more severe consequences, beginning with providing students with a transfer option and ending with closure or conversion into a charter or private school.

Accountability systems can be designed with either stronger or weaker linkages between outcomes and incentives or consequences. Some argue that accountability systems with low stakes for educators will not induce them to improve educational practice, and push for strong consequences associated with measured performance. However, the problem with high-stakes accountability is that the objective metrics are typically incomplete descriptions of performance. Schools are trying to accomplish many objectives—higher student achievement on certain tests, but also achievement in those areas that may not be well-captured by performance on standardized tests, performance in other academic areas that do not appear on the accountability test, and more abstract goals such as critical thinking, open-mindedness, maturity, and citizenship. When faced with strong incentives to concentrate on some metrics but not on others, schools might be expected to focus on short-run gains in what is being measured—sometimes obtained through strategic behavior such as “teaching to the test”—at the expense of long-run skill acquisition. Moreover, even low-stakes accountability systems that are based exclusively on information can have high stakes for educators if stakeholders respond to that information (Figlio and Lucas 2004; Figlio and Kenny 2009).

With this conceptual framework in mind, school accountability has often been studied as an application of a multitask moral hazard model where performance measures are used in place of a true objective that cannot be observed directly (Holmstrom and Milgrom 1991; Baker 1992; MacLeod 2003). In the Holmstrom and Milgrom (1991) theoretical analysis of these models, they use the example of teachers teaching basic skills and higher-order thinking, where the latter cannot be measured. The key insight from these models is that the optimal strength of

performance incentives is increasing in the correlation between a performance metric (say, high-stakes tests) and the true objective (say, developing broader capacities, or perhaps earnings). Crucially, it is the correlation *at the margin* that matters (Hout and Elliott 2011). When schools face pressure to raise test scores, and they take action, what is the effect of those actions on the long-run outcomes that are the true objective of schooling? When the correlation between test score gains and gains in long-run outcomes is weak, low-powered incentives, or even no incentives at all may be preferable.

Moreover, an inherent tension also arises between using achievement tests both as a diagnostic tool and also as a high-stakes performance measure (Neal 2013). This follows from what is known colloquially as Campbell's law (1976)—“the more any quantitative social science indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.”

This discussion suggests that as long as what is being measured is only a proxy for the truly desired outcome, the effects of accountability efforts are theoretically ambiguous. Raising the stakes magnifies the impact of accountability on behavior, but whether students are helped or harmed overall by the changes in behavior is ultimately an empirical question.

Evidence on Accountability in K–12 Education

Reactions to Accountability

The evidence suggests that families respond strongly to provision of information about the K–12 education system, even in the absence of explicit stakes. For example, differences in test scores are capitalized into housing markets. In an early study of this effect, Black (1999) looked at Massachusetts houses that were located close to the boundary between school districts and found that homeowners were willing to pay 2.5 percent more for a 5 percent increase in test scores. For a literature review of later studies finding a broadly similar result, see Black and Machin (2011). Using a similar approach of looking at house location and school zone boundaries with Florida data, Figlio and Lucas (2004) find that the grades given to schools by the state of Florida also affect real estate values (and the effect of grades is in addition to the effect of any change in test scores). In another study of Florida data, Figlio and Kenny (2009) took advantage of an administrative shift in how Florida graded its schools to show that voluntary contributions to schools (typically made through parent–teacher organizations) rise and fall based on accountability measures. Hastings and Weinstein (2008) show that when parents are making choices between schools in the Charlotte-Mecklenburg school district, when they receive new information about test scores in schools, they are more likely to choose the schools with higher test scores.

Those who work in schools respond to accountability ratings, too. Several case studies have found that principals and teachers perceive that their job security is

tied to their school's accountability rating; for example, Toenjes and Garst (2000) discuss this connection for Texas schools and districts; the collection of essays in Evers and Walberg (2002) includes a comparison of accountability systems in Texas, Florida, and California; Lemons, Luschei, and Siskin (2003) conduct case studies of six high schools in two states; and Mintrop and Trujillo (2005) compare three medium-sized states (Kentucky, Maryland, and North Carolina), four larger ones (California, Florida, New York, and Texas), and two large districts (Chicago and Philadelphia).

State and federal accountability policies typically combine information provision with performance incentives in a variety of ways, making it difficult to distinguish one from the other. However, a lot of evidence suggests that K–12 public schools do respond to accountability pressure. Several studies of state and federal accountability systems have found modest positive impacts on low-stakes test scores in reading and math, including in single-state or single-district studies that use differences in program rules for identification (Chiang 2009, Figlio and Rouse 2006, and Greene, Winters, and Forster 2004, all in Florida; Rockoff and Turner 2010, in New York City; Jacob 2005, in Chicago; and Ladd 1999, in Dallas) and overseas (Allen and Burgess 2012, in the United Kingdom). These studies tend to find larger effects for math than for reading, though in some cases (for example, Jacob 2005) the estimated effects are larger for reading. A typical finding suggests that accountability boosts math test scores in the lowest-performing schools by about one-tenth of a standard deviation relative to those in higher-performing schools, though the studies exhibit a wide range of impacts.

Researchers have tended to focus on low-stakes assessments because of concerns that gains on high-stakes tests reflect strategic responses to accountability pressures rather than genuine improvements. Winters, Trivitt, and Greene (2010) demonstrate that Florida's accountability system improved outcomes not just in math and reading, but also had spillover effects into science, even though there were no stakes attached to science performance at the time. Rouse, Hannaway, Goldhaber, and Figlio (2013) find that schools facing accountability pressure do change their practice: for example, such schools reorganize the school day and the learning environment to focus on low-performing students and lengthen the amount of instruction time, while also increasing resources available to teachers. They show that these changes in policies and practices account for a substantial fraction of the test score improvement. Dee and Jacob (2011) find that practices installed as a result of the No Child Left Behind legislation increased student achievement on the low-stakes National Assessment of Educational Progress.

We are aware of only one study that investigates the impact of K–12 school accountability on long-run outcomes. Deming, Cohodes, Jennings, and Jencks (forthcoming) find that accountability pressure in Texas high schools led to increases in college attainment and earnings for low-scoring students in low-scoring schools. However, they also found some evidence of *negative* impacts for other students, which they argue arose from schools' strategic responses to the rules around student testing exemptions. As we discuss in the next section, there are

many examples of so-called “strategic” responses to school accountability. This foreshadows a key design challenge for accountability in US higher education—schools typically respond strongly to performance incentives, but not always in the ways that the advocates of such incentives would like.

Critiques of Accountability: Unclear Information and Strategic Responses

One criticism of “report card” accountability is that governments are not very good at providing information in an easily digestible format. Yet here, for better or worse, the private market has already stepped in. Websites such as *greatschools.org*, *schoolgrades.org*, *k12.niche.com*, and *schooldigger.com* make their living by translating publicly available information about school quality into user-friendly formats. However, a more fundamental question is the extent to which this information captures school quality accurately or comprehensively. Public provision of inaccurate or noisy measures of quality cannot be expected to improve student outcomes in any meaningful way.

Economists often focus on wages as an omnibus measure of impact. But drawing causal connections between test scores (or graduation rates) today and wages in the future is extremely difficult. A broader concern, however, is that school quality is subjective and multidimensional. As a result, the benchmark measures chosen for K–12 or postsecondary accountability, such as levels or gains or pass rates on certain tests, are always going to be incomplete proxies for the overall goals of education.

In this situation, one should be concerned that reducing the accounting of school quality to a small number of conveniently measurable outcomes narrows the focus of actors within an educational system. Indeed many studies have found that schools facing accountability pressure narrow their curriculum or instructional practices at the expense of nontested groups or subjects (Stecher, Barron, Chun, and Ross 2000; Diamond and Spillane 2004; Booher-Jennings 2005; Hamilton, Berends, and Stecher 2005; Diamond 2007; Ladd and Lauen 2010; Reback 2008; Neal and Schanzenbach 2010; Özek 2012; Reback, Rockoff, and Schwartz 2014).

Narrowing the curriculum is not necessarily a bad thing. Yet some strategic responses to accountability are harder to justify. Figlio and Winicki (2005) demonstrate that Virginia schools subject to accountability pressure strategically raise the calorie content of meals on test days, and find suggestive evidence that schools that use this approach see a larger rise in high-stakes pass rates. The fact that schools react in this manner, and that their reactions lead to improved measured outcomes, shows one way in which the reporting of measures of school quality can help to undermine their validity. Other studies have suggested responses that are potentially more insidious, such as strategic reclassification of students into disability categories (Deere and Strayer 2001; Cullen and Reback 2006; Figlio and Getzler 2006; Deming et al. forthcoming), using disciplinary procedures to suspend low-performing students from school when the tests are given (Figlio 2006), and outright teacher cheating (Jacob and Levitt 2003).

Lessons about Accountability from the K–12 Experience

Four lessons emerge from the existing literature on performance measurement and accountability in K–12 education.¹

First, when public reporting and rewards and sanctions are tied to specific measures, as is often the case, organizations will seek to maximize short-run performance on those measures at the potential expense of other outcomes of interest.

Second, design details of accountability metrics strongly influence organizational behavior. For example, consider the difference that arises if an assessment measure is based on the share of students who exceed a certain proficiency target, or if it is based on the value-added gains that students made from their scores in the previous year. When K–12 schools are assessed based on proficiency targets, they have a strong incentive to focus on “bubble” students who are near the threshold, or on other students more likely to “count” for accountability (Neal 2010; Neal and Schanzenbach 2010; Figlio and Loeb 2011; Özek 2012; Figlio and Ladd 2015).

While value-added metrics reduce incentives to focus on particular students (after all, it is much harder to target students with high potential *gains*), they may introduce additional scope for distortions. For example, Macartney (2016) finds that schools and teachers in North Carolina responded to value-added performance targets by *reducing* effort in earlier periods—the so-called “ratchet effect.” In addition, yearly fluctuation in test scores may make value-added metrics quite noisy and difficult for families to understand and use (Kane and Staiger 2002; Chay, McEwan, and Urquiola 2005). Indeed, Brehm, Imberman, and Lovenheim (2015) show using data from the Houston school district that student test score gains were so noisy that incentives based on value-added measures failed to elicit any performance response from teachers.

Third, the scope for strategic responses to accountability increases with the number and complexity of high-stakes metrics. A typical K–12 accountability system employs tests in multiple subjects, across multiple (but not all) grades and student subgroups, and includes a frequently byzantine set of exemptions and secondary metrics and assessments. Every additional layer of complexity introduces more opportunities for strategic behavior.

¹Similar lessons arise from evidence on performance incentives in other settings. Institutions do respond to the information embedded in accountability systems, but not always in socially desired ways. For examples of some positive reactions, see Jin and Leslie (2003) on the reduction in food-related hospitalizations after Los Angeles County posted restaurant hygiene grade cards, and Bennear and Olmstead (2008) on how utilities that were required to disclose customer confidence reports reduced their health violations. For examples of some accountability systems with mixed results, Heckman, Heinrich, and Smith (1997, 2002) analyze performance standards for Job Training Partnership Act (JTPA) centers; Cutler, Huckman, and Landrum (2004) show that cardiac surgery report cards in New York led to patient selection and subsequent improvement of poorly performing hospitals, although Dranove, Kesser, McClellan, and Satterthwaite (2003) find that these report cards (and those in Pennsylvania) also resulted in higher levels of resource use and reduced health outcomes; and Lu (2012) demonstrates that the Nursing Home Quality Initiative led to improvements in reported measures of quality but deterioration in unreported areas.

Fourth, accountability works best at improving results at the bottom of the distribution. The balance of the evidence from studies of accountability, either across states or within them, suggests larger gains for low-income, minority, and low-achieving students (Carnoy and Loeb 2002; Dee and Jacob 2011; Lauen and Gaddis 2012; Deming et al. forthcoming). Additionally, many of the studies that find positive impacts of accountability pressure compare schools on either side of a cutoff that defines a “failing” grade (Figlio and Rouse 2006; Chiang 2009; Allen and Burgess 2012; Rouse et al. 2013; Reback, Rockoff, and Schwartz 2014). These schools are by definition among the lowest-achieving in the state.²

Researchers rarely find much of a response to accountability pressure in higher-performing schools. This may be because the lowest performance threshold is the most salient to educators and households, and confers the greatest stigma. Alternatively, families with higher socioeconomic status may monitor schools more closely, leading them to rely less on the public signal sent by an accountability rating. Thus, external accountability works best when institutions would not otherwise face strong internal or community pressures to improve. Families can hold schools accountable by monitoring school performance, and such monitoring may be more intense among affluent households (Ferreyra and Liang 2012). Moreover, affluent families may be more likely to sort across neighborhoods in response to perceived changes in school quality, which also places accountability pressure on schools (Bayer, Ferreira, and McMillan 2007). In contrast, schools serving disadvantaged populations may face less parental pressure or lack the capacity for self-monitoring, making external accountability relatively more important.

Accountability in US Higher Education

In September 2015, the US government released its “College Scorecard,” which includes data from the US Department of the Treasury and Internal Revenue Service, as well as US Department of Education records covering over 7,000 colleges and universities (available at <http://collegescorecard.ed.gov>). The College Scorecard includes information on average college costs, overall and by family income levels; typical student debt loads, fraction of students receiving federal loans, and fraction of students making good progress in paying down their debt; graduation and one-year retention rates; median earnings of students ten years after entering the college; as well as student body characteristics (racial/ethnic breakdown, socioeconomic diversity, and college entry exam scores). These data provide, for the first time, a remarkable wealth of information on not just student body

²Sometimes, perhaps due to incentives to boost performance of marginal students, the lowest-performing students in low-achieving schools do not appear to benefit from accountability, even when more marginal students in their low-performing schools do (for example, Deming forthcoming, in Texas), perhaps due to the types of strategic behaviors such as focusing on “bubble kids” (Neal and Schanzenbach 2010) described above.

characteristics—a mainstay of private college ratings—but also some downstream outcomes of colleges.

Making the College Scorecard data available is a public service. However, the variables that are included and excluded send signals to the general public about what is valued and what is not, which in turn raises underlying questions: What are the desired outcomes of postsecondary education? And what are the likely outcomes of making data available and/or constructing rankings based on postsecondary data? While the federal government assiduously avoided constructing explicit college ratings with these data, it is certainly possible to construct ratings using them, and numerous organizations have done so.

An array of studies have found that higher education institutions respond strategically to privately produced institutional rankings, most notably those produced by *U.S. News and World Report*. For example, Monks and Ehrenberg (1999) look at how year-to-year changes in USNWR rankings from the late 1980s through the 1990s influenced the admissions outcomes and pricing policies at selective colleges, while Meredith (2004) finds similar results while studying a broader number of schools. Several studies suggest that the visibility and salience of the rankings are important, even aside from their quality. For example, Luca and Smith (2013) find that a one-rank improvement in the USNWR rankings leads to a one-percentage-point increase in applications. However, this effect only appears when the ranked institutions are listed numerically, and disappears when they are listed alphabetically. Similarly, Bowman and Bastedo (2009), find that being on the front page of the USNWR rankings, or not, has an effect on admissions. Bastedo and Bowman (2010) find that the perceptions of quality expressed by senior administrators at peer institutions are affected by the USNWR ratings, and Espeland and Sauder (2007) document with interview data how the rankings cause laws schools to change their behavior and expectations.

The Case for Information without Rankings

Although students and university administrators undoubtedly pay attention to college rankings, the impact of the rankings on institutional behavior, prices, and student outcomes is much less clear. In some ways, this combination of information without consequential accountability may be a healthy situation. Compared to their K–12 counterparts, US colleges and universities have broader purposes and serve a greater variety of students. This diversity across institutions greatly increases the degree of difficulty in designing an effective accountability system, because the benchmarks are harder to agree upon and the scope for strategic responses is much greater.

While K–12 schools are mostly required to take all comers, postsecondary institutions choose which students to admit—and even open enrollment institutions engage in subtle forms of selection. Public K–12 institutions have clear and well-defined missions and offer a “standard” curriculum. In contrast, higher education institutions decide which programs to offer and differ greatly in their stated institutional missions. Colleges and universities also operate in very different markets, ranging

from open-access community colleges with a mandate to serve the local economy to elite institutions that compete for the best students on a global scale.

For these reasons, trying to rank colleges and universities on a few common standards may not make sense. To give just one example, the College Scorecard lists both Boston University and the New England Conservatory of Music as having an average annual cost (defined as the average price net of all financial aid) of around \$35,000. Yet the average salary for Boston University graduates ten years later is more than double (\$60,600 vs. \$29,500). Are we comfortable rating colleges according to a financial benefit–cost calculation that will undoubtedly penalize students who self-select into lower-earning fields of study?

One potential consequence of a college rating system is that selective institutions might become even more stratified. MacLeod and Urquiola (2015) show that reputational incentives lead to stratification even in the absence of direct peer effects. In their model, ability is imperfectly observed and employers use college reputation (defined as the average skill of its graduates) as a signal of worker skill. As a result, students endogenously prefer better peers because of the signal that college reputation sends to the market. Importantly, stratification increases study effort prior to college admission and reduces study effort *in college*. Thus by selecting high-ability students at entry, colleges can have “elite” reputations without necessarily having high value-added to their graduates.

Between the difficulties of ranking institutions of higher education and the expectation that sorting will occur between them, an option is to forgo explicit stakes altogether, focusing instead on providing transparent and easily digestible information about school characteristics and performance, and letting consumers use it as they see fit. This is the College Scorecard approach, and it has much appeal. However, one lesson from K–12 accountability is that information alone can be a powerful driver of decision-making. Thus even if we decide that “report card” accountability for higher education is sufficient, we must think carefully about what information to provide, and in what way.

Targets and Tradeoffs for Accountability: Graduation Rates, Debt, Employment Outcomes, Direct Exams

In thinking about the impact of consequential accountability for US higher education, it is useful to start first with the variables actually used by the College Scorecard.

First, suppose that institutional ratings were based on graduation rates, borrowing rates, default rates, and/or borrowing intensity. These variables can, of course, be affected by institutional quality or policy decisions such as generosity of financial aid. But measured “success” in such a situation is surely also determined by student quality at admission. Institutions that serve primarily disadvantaged or first-generation college students will—all else equal—have lower graduation rates and higher borrowing rates. As a consequence, institutions that have missions to educate larger numbers of first-generation and disadvantaged students will look less attractive by these criteria. Colleges might alter admission criteria in order to

improve the likelihood that admitted students will be able to succeed and pay for college without much borrowing.

In principle, one could “risk adjust” performance standards to reflect pre-existing differences in the likelihood of student success. This can reduce some aspects of selection, but as with “value-added” approaches in K–12 education, risk adjustment can also increase measurement error and reduce transparency and public confidence in the rating system. Barnow and Heinrich (2010) discuss the benefits and costs of such risk adjustment in a variety of settings. In higher education, there are large differences in students’ prior preparation even within open access institutions (Kurlaender, Carrell, and Jackson 2016). As a result, failure to adjust for differential selection could be quite problematic.

Moreover, the data reported in the College Scorecard are calculated based on different populations for each outcome. For instance, while college graduation rates are calculated for all students, average costs of attendance and subsequent earnings are calculated only among federal financial aid recipients. Thus, outcome data are missing for some types of students and not others, making some type of risk adjustment extremely important for apples-to-apples comparisons across institutions.

Evaluating institutions on the basis of employment and earnings outcomes involves many of the same complications as other performance metrics. However, an additional complication comes from the wide variety in average compensation by field of study. Four-year college graduates with the highest-paying majors earn two-and-a-half times on average what the four-year college graduates with the lowest-paying majors earn (Hershbein and Kearney 2014). Majors that prepare students to work with children (like early childhood education and elementary education) or provide community and counseling services (like family sciences, social work, and theology) have the lowest average earnings. Evaluating institutions on one dimension like earnings could lead to reductions in opportunities to prepare for fields that are socially desirable but not financially lucrative; this is one example of how accountability can exacerbate the multitasking problem in higher education. In addition, Oreopoulos and Salvanes (2011) document a number of nonpecuniary benefits of postsecondary study that are not captured by labor market outcomes. It is not difficult to imagine that some colleges provide great value-added for “nonmarket outcomes” that do not show up on the balance sheet.

Another limitation of using employment outcomes for accountability is the long time horizon required to measure post-college earnings. The College Scorecard measures earnings ten years after initial enrollment. If colleges are evaluated based on earnings in a student’s late 20s, a relatively easy way to “game the system” is to emphasize fields of study where early career earnings are high and graduate education is uncommon—the Scorecard excludes individuals known to be enrolled in school at the point of measurement so graduate students ten years out don’t help the school’s ratings—or by counseling students into higher-earning options. On the other hand, a substantially longer time horizon means that the information on earnings is what happened to those who enrolled more than a decade earlier, and colleges are unlikely to alter their behavior based on predictions of outcomes far in the future.

A final possibility is to hold postsecondary institutions accountable for learning outcomes directly, using assessments such as the National Survey of Student Engagement or the Collegiate Learning Assessment (for example, Arum and Roksa 2010). This approach presents many opportunities for institutional strategic behavior observed at the K–12 level, both in terms of emphasizing the types of skills that are more likely to be represented on the assessment as well as in terms of selecting which students enroll and actually take the assessment. Moreover, while tools like the Collegiate Learning Assessment are surely valuable indicators of one aspect of student learning growth over the course, they do not reflect the wide range of objectives of postsecondary institutions or the considerable heterogeneity in the purposes of these institutions. It is possible to construct field-specific exit exams that would at least present the opportunity to capture one aspect of skill; MacLeod, Riehl, Saavedra, and Urquiola (2015) present evidence from Colombia that the rollout of a field-specific college exit exam reduced some of the labor market returns to college reputation.³ But carrying out this form of exit exam is extremely expensive, and the introduction of such an exam brings with it the risk of new manipulative behaviors on the part of educational institutions, along with the challenges associated with measuring institutional value added, which are surely more difficult in the postsecondary setting.

Some Design Principles for Accountability in Higher Education

Some of the adjustments and tradeoffs from greater accountability in higher education may be welcome. After all, if students who are undecided about majors get a nudge toward a choice that pays better, or if schools put more emphasis on a high graduation rate and a lower debt burden, such steps may overall be beneficial. Indeed since all college students are paying customers—both directly and indirectly through the opportunity cost of foregone immediate earnings—we might expect accountability to have a larger impact in higher education than in K–12. Here, we draw on insights from economic theory and from lessons learned in K–12 education to lay out some design principles for accountability in US higher education.

A first design principle is that a college rating system should be kept as simple as possible to reduce the scope for strategic responses. While some risk adjustment is probably necessary, it should be as transparent as possible to facilitate consumer choice. Rather than constructing college “value added” through regression adjustment, a simpler alternative is to construct groups of postsecondary institutions that represent likely choice sets for certain types of students. Equivalence classes could be created based on geographical proximity and measures of selectivity, or alternatively they could be constructed empirically using overlap in actual students’ choice sets (for example, Avery, Glickman, Hoxby, and Metrick 2012).

A second design principle is to target the postsecondary institutions that are least likely to respond to market forces in the absence of accountability. In the K–12

³Hoekstra (2009) and others demonstrate that there exists a labor market return to higher education institutional reputation in the United States.

setting, most of the benefits of accountability come from pressure on educators to avoid a failing grade, perhaps because families with higher socioeconomic status monitor schools more closely and are more likely to “vote with their feet.” Likewise, elite colleges already compete fiercely for students, and a government rating is unlikely to change their incentives much. Thus one idea is to focus on certifying a minimum standard of quality, rather than assigning grades or ratings to institutions all along the spectrum. Similar to health inspections or the consumer drug approval process, the job of a higher education accountability system could be to certify that schools are good enough to receive public support.

Public certification of postsecondary institutions already exists in the form of accreditation. The US Department of Education keeps a list of regional and national accreditors, and in principle institutions must be approved by an accreditor’s regular inspections to distribute federal financial aid. Yet in practice, accreditors—who are paid by the institutions themselves—appear to be ineffectual at best, much like the role of credit rating agencies during the recent financial crisis. As a case in point, the Accrediting Council for Independent Colleges and Schools (ACICS) has come under considerable scrutiny for continuing to accredit branches of Corinthian Colleges right up until the company’s collapse in April 2015 amid allegations of fraud and financial misconduct.

While we are unaware of well-identified studies of the consequences of independent accreditation in the higher education sector, Hussain (2015) demonstrates that in the K–12 sector in the United Kingdom, inspectorate systems led to measurable and lasting improvements in student outcomes. One possible approach is to design an inspectorate system that is “turned on” when an institution falls below quantitative benchmarks. While school inspections are resource-intensive, targeting toward the lowest performers would help to limit the cost of such a program. Duflo, Greenstone, Pande, and Ryan (2013) present experimental evidence from environmental inspections of industrial plants in India. They find that regulation works much better when auditors are randomly assigned and centrally compensated, rather than chosen and paid by firms themselves. Similar reforms to accreditation might have sizeable benefits in terms of improved higher education outcomes.

Another way that higher education accountability can target the lowest performing institutions is by setting a relatively low bar for performance yet enforcing it vigorously. The federal Gainful Employment regulations that went into effect in 2015 are one—albeit imperfect—example. The purpose of the Gainful Employment regulations is to link the costs and benefits of postsecondary programs explicitly (US Department of Education 2015), and while Gainful Employment regulates on debt burden alone, it is still a step in this direction. The rules specify that, on average, graduates of nearly all for-profit programs (along with certificate programs at not-for-profit and public institutions) must have an annual loan payment that does not exceed 20 percent of discretionary income or 8 percent of total earnings. The penalty for repeatedly falling below this debt-to-earnings threshold is the withdrawal of eligibility for that institution to disburse federal Title IV financial aid.

This bar seems relatively low, yet the Department of Education estimated that 840,000 students were enrolled in programs that would not have met the standard in 2013. While this represents fewer than 5 percent of all US postsecondary enrollment, more than 99 percent of the students were concentrated within a small number of for-profit programs. The combination of high prices and low labor market returns is unique to the for-profit higher education sector, making it a prime target for increased accountability. Looney and Yannelis (2015) show that for-profit institutions are responsible for a disproportionate share of the increase in student debt and loan defaults since 2000, and Deming et al. (2016) use a resume audit experiment to show that employers are less likely to express interest in a resume with a degree from a for-profit institution of higher education compared to identical resumes with degrees from public institutions.

The design of Gainful Employment is simple and straightforward, and the regulation successfully concentrates on the worst offenders. However, a legitimate criticism is that it unfairly targets the for-profit sector and leaves poorly performing public institutions untouched.⁴

Ideally, regulations like Gainful Employment would focus on institutions that rely heavily on public subsidies—regardless of their for-profit or public status. At present, for-profit institutions derive about 75 percent of their revenue from federal Title IV Pell Grants and Stafford Loans, which are disbursed to eligible students based on financial need (Deming, Goldin, and Katz 2012). A federal regulation known as the 90/10 rule prohibits these colleges from deriving more than 90 percent of revenue from Title IV aid. Yet the largest for-profit colleges bump right up against this 90 percent cap.⁵ This dependence on taxpayer largesse, more than for-profit status, justifies tighter regulation. Many smaller for-profit institutions attract paying customers without needing federal financial aid subsidies, and these schools are rightly free from the Gainful Employment regulations (for example, Cellini and Goldin 2014).

In most states, community colleges and less-selective four-year publics are also heavily subsidized by taxpayers. Tuition is kept much lower than the resource cost of college, and is in some cases close to zero after accounting for federal financial aid.

⁴The Gainful Employment program focuses only on for-profits and certificate programs in nonprofit and public institutions. This targeting was partly a regulatory necessity—the phrase “gainful employment” originates from language in the Higher Education Act of 1965 that specifies which institutions are allowed to distribute Title IV aid—but was also deliberately aimed at the for-profit sector. For-profits have criticized the Gainful Employment regulations for unfairly targeting the sector. In 2012, a for-profit college trade group sued, and the initial regulation from the US Department of Education, which included a rule about loan repayment rates in addition to the debt-to-earnings ratio, was struck down in federal district court. The follow-up effort, which eliminated repayment rates as an accountability metric, was upheld in May 2015. This court decision set the stage for Gainful Employment rules to become law in July 2015, although the ruling remains legally tenuous.

⁵In addition, the large for-profits engage in strategic behavior that seems aimed at maximizing loan support, such as targeted recruitment of GI bill-subsidized military students. Military students receive higher education subsidies from the GI bill, which is federal aid to students in higher education that does not fall under Title IV. Thus every \$1 of GI bill subsidies allows schools to bring in another \$9 of Title IV aid while remaining under the 90 percent revenue cap.

Schools that depend more on taxpayer support should justifiably be targeted with increased government regulation, even if they are public institutions. Concretely, one could design an accountability system where regulatory control is increasing in the share of institutional revenue that comes from public sources.

Thus a third principle, following from the discussion above, is that accountability in higher education should be designed to ensure that both students and postsecondary institutions have some “skin in the game.” Here there is no exact parallel with K–12 schooling, because most college students and few primary and secondary school students pay for their education. However, “skin in the game” can be a guiding principle for regulators in thinking about how much control to exert over postsecondary institutions. Colleges that can attract full-paying customers—either out-of-state students or students who do not qualify for financial aid—have implicitly survived a market test and should be allowed to operate more freely. This principle does not mean that public institutions cannot be heavily subsidized, but it does suggest that scrutiny should be greater when taxpayers are footing more of the bill.

A more direct approach is risk-sharing, where institutions would be responsible for paying a share of student loans that subsequently end up in default. As with all accountability metrics, risk-sharing programs would probably lead to lower lending, but also with some potential for strategic responses. Institutions would be more likely to enroll students whom they suspect will stand the best chance of completing college and repaying their student loans. Institutions might also offer fewer programs in professions with high social value but low downstream income potential.

Conclusion

The rationale for increased accountability in the higher education sector is clear. However, designing a well-functioning accountability system is extremely difficult. The experience from accountability in K–12 education and other industries demonstrates that “what gets measured gets done,” in both socially desirable and undesirable ways. Also, figuring out what *should* get measured and what *should* get done is no easy matter. The outcomes desired by parents and students might differ from the outcomes chosen by policymakers, and any one-size-fits-all solution will not do full justice to the multidimensional nature of higher education.

One main lesson we take from the research evidence is that accountability is likely to be most important in the education markets that are the least competitive. At the K–12 level, accountability works best in low-performing schools with weak systems of support, and when students have relatively few options other than their local public school. Similarly, we suspect that accountability for selective colleges will have little impact, because both elite colleges and the students who attend them already have plenty of “skin in the game.” Just to be clear, this does not mean that we believe all students at such institutions receive a high-quality and cost-effective

education. Rather, we believe that additional accountability measures aren't likely to lead to improvements among schools that are already facing other kinds of pressure, although they may respond strategically to improve their position in the ratings hierarchy.

This same logic also applies to less-selective institutions of higher education that can attract paying customers with only modest help from public funds or federal financial aid. In contrast, for postsecondary institutions that are heavily dependent on taxpayer support, or that have a poor record on metrics like graduation rates, an accountability system with explicit consequences could improve student outcomes.

If performance measures work, they will provoke a mix of real improvement and strategic responses. Thus, an ebb-and-flow of defining accountability, backing away, and then redefining accountability and backing away again, is to be expected. As this process evolves, the ongoing challenge is to maximize the benefits of accountability while minimizing its unintended side effects. In higher education, this may be the time for state-level policy experimentation: if different states try different forms of accountability for their higher education institutions and programs, we will have the opportunity to learn more about which approaches to accountability in the higher education sector yield the greatest net benefits.

■ *The authors are grateful to Mark Gertler, Claudia Goldin, Gordon Hanson, Larry Katz, Jordan Matsudaira, Enrico Moretti, Timothy Taylor, and Miguel Urquiola for helpful comments and suggestions, though all errors are their own.*

References

- Allen, Rebecca, and Simon Burgess.** 2012. "How Should We Treat Under-performing Schools? A Regression Discontinuity Analysis of School Inspections in England." Working Paper no. 12/287, Center for Market and Public Organization, University of Bristol.
- Arum, Richard, and Josipa Roksa.** 2010. *Academically Adrift: Limited Learning on College Campuses*. University of Chicago Press.
- Avery, Christopher N., Mark E. Glickman, Caroline M. Hoxby, and Andrew Metrick.** 2013. "A Revealed Preference Ranking of American Colleges and Universities." *Quarterly Journal of Economics* 128(1): 425–67.
- Avery, Christopher, and Sarah Turner.** 2012. "Student Loans: Do College Students Borrow Too Much—Or Not Enough?" *Journal of Economic Perspectives* 26(1): 165–92.
- Baker, George P.** 1992. "Incentive Contracts and Performance Measurement." *Journal of Political Economy* 100(3): 598–614.
- Barnow, Burt S., and Carolyn J. Heinrich.** 2010. "One Standard Fits All? The Pros and Cons of Performance Standard Adjustments." *Public Administration Review* 70(1): 60–71.
- Bastedo, Michael N., and Nicholas A. Bowman.** 2010. "U.S. News and World Report College Rankings: Modeling Institutional Effects on Organizational Reputation." *American Journal of Education* 116(2): 163–83.
- Bayer, Patrick, Fernando Ferreira, and Robert McMillan.** 2007. "A Unified Framework for

Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy* 115(4): 588–638.

Benneer, Lori S., and Sheila M. Olmstead. 2008. "The Impacts of the 'Right to Know': Information Disclosure and the Violation of Drinking Water Standards." *Journal of Environmental Economics and Management* 56(2): 117–30.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics* 114(2): 577–99.

Black, Sandra E., and Stephen Machin. 2011. "Housing Valuation and School Performance." Chap. 10 in *Handbook in Economics of Education*, vol. 3, edited by Eric Hanushek, Stephen Machin, and Ludger Woessmann. Elsevier.

Booher-Jennings, Jennifer. 2005. "Below the Bubble: 'Educational Triage' and the Texas Accountability System." *American Educational Research Journal* 42(2): 231–68.

Bowman, Nicholas A., and Michael N. Bastedo. 2009. "Getting on the Front Page: Organizational Reputation, Status Signals, and the Impact of U.S. News and World Report on Student Decisions." *Research in Higher Education* 50(5): 415–36.

Brehm, Margaret, Scott A. Imberman, and Michael F. Lovenheim. 2015. "Achievement Effects of Individual Performance Incentives in a Teacher Merit Pay Tournament." NBER Paper 21598.

Campbell, Donald T. 1976. "Assessing the Impact of Planned Social Change." Occasional Paper 8, December. (Reprinted February 2011 in the *Journal of MultiDisciplinary Evaluation*, vol. 7, no. 15.)

Carnoy, Martin, and Susanna Loeb. 2002. "Does External Accountability Affect Student Outcomes? A Cross-State Analysis." *Educational Evaluation and Policy Analysis* 24(4): 305–31.

Chay, Kenneth Y., Patrick J. McEwan, and Miguel Urquiola. 2005. "The Central Role of Noise in Evaluating Interventions That Use Test Score to Rank Schools." *American Economic Review* 95(4): 1237–58.

Cellini, Stephanie Riegg, and Claudia Goldin. 2014. "Does Federal Student Aid Raise Tuition? New Evidence on For-Profit Colleges." *American Economic Journal: Economic Policy* 6(4): 214–206.

Chiang, Hanley. 2009. "How Accountability Pressure on Failing Schools Affects Student Achievement." *Journal of Public Economics* 93(9–10): 1045–57.

Cullen, Julie Berry, and Randall Reback. 2006. "Tinkering toward Accolades: School Gaming under a Performance Accountability System." In *Advances in Applied Microeconomics*, vol. 14, edited by Timothy J. Gronberg and Dennis W. Jansen, 1–34. Elsevier.

Cutler, David M., Robert S. Huckman, and Mary Beth Landrum. 2004. "The Role of Information in Medical Markets: An Analysis of Publicly Reported Outcomes in Cardiac Surgery." *American Economic Review* 94(2): 342–46.

Dee, Thomas S., and Brian Jacob. 2011. "The Impact of No Child Left Behind on Student Achievement." *Journal of Policy Analysis and Management* 30(3): 418–46.

Deere, Donald, and Wayne Strayer. 2001. "Putting Schools to the Test: School Accountability, Incentives and Behavior." Unpublished paper, Texas A&M University.

Deming, David J., Sarah Cohodes, Jennifer Jennings, and Christopher Jencks. Forthcoming. "School Accountability, Postsecondary Attainment and Earnings." *Review of Economics and Statistics*.

Deming, David J., Claudia Goldin, and Lawrence F. Katz. 2012. "The For-Profit Postsecondary School Sector: Nimble Critters or Agile Predators?" *Journal of Economic Perspectives* 26(1): 139–64.

Deming, David J., Noam Yuchtman, Amira Abulafi, Claudia Goldin, and Lawrence F. Katz. 2016. "The Value of Postsecondary Credentials in the Labor Market: An Experimental Study." *American Economic Review* 106(3): 778–806.

Diamond, John B. 2007. "Where the Rubber Meets the Road: Rethinking the Connection between High-Stakes Testing Policy and Classroom Instruction." *Sociology of Education* 80(4): 285–313.

Diamond, John B., and James Spillane. 2004. "High-Stakes Accountability in Urban Elementary Schools: Challenging or Reproducing Inequality?" *Teachers College Record* 106(6): 1145–76.

Dranove, David, Daniel Kessler, Mark McClellan, and Mark Satterthwaite. 2003. "Is More Information Better? The Effects of 'Report Cards' on Health Care Providers." *Journal of Political Economy* 111(3): 555–88.

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2013. "Truth-Telling by Third-Party Auditors and the Response of the Polluting Firms: Experimental Evidence from India." *Quarterly Journal of Economics* 128(4): 1449–98.

Espeland, Wendy Nelson, and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113(1): 1–40.

Evers, William, and Herbert Walberg. 2002. *School Accountability*. Palo Alto, CA: Hoover Press.

Ferreira, Maria Marta, and Pierre Jinghong Liang. 2012. "Information Asymmetry and Equilibrium Monitoring in Education." *Journal of Public Economics* 96(1–2): 237–54.

Figlio, David N. 2006. "Testing, Crime and Punishment." *Journal of Public Economics* 90(4–5):

837–51.

Figlio, David N., and Lawrence S. Getzler. 2006. “Accountability, Ability and Disability: Gaming the System?” In *Advances in Applied Microeconomics*, Vol 14: *Improving School Accountability* edited by Timothy J. Gronberg and Dennis W. Jansen, 35–39. Emerald.

Figlio, David N., and Lawrence W. Kenny. 2009. “Public Sector Performance Measurement and Stakeholder Support.” *Journal of Public Economics* 93(9–10): 1069–77.

Figlio, David N., and Helen F. Ladd. 2015. “School Accountability and Student Achievement.” In *Handbook of Research in Education Finance and Policy*, 2nd edition, edited by Helen F. Ladd and Margaret Goertz, pp. 194–210. AAFP and Routledge.

Figlio, David, and Susanna Loeb. 2011. “School Accountability.” In *Handbook in Economics of Education*, vol. 3, edited by Eric Hanushek, Stephen Machin, and Ludger Woessmann, 383–421. Elsevier.

Figlio, David N., and Maurice E. Lucas. 2004. “What’s in a Grade? School Report Cards and the Housing Market.” *American Economic Review* 94(3): 591–604.

Figlio, David N., and Cecilia Elena Rouse. 2006. “Do Accountability and Voucher Threats Improve Low-Performing Schools?” *Journal of Public Economics* 90(1–2): 239–55.

Figlio, David N., and Joshua Winicki. 2005. “Food for Thought: The Effects of School Accountability Plans on School Nutrition.” *Journal of Public Economics* 89(2–3): 381–94.

Friedman, Milton. 1962. *Capitalism and Freedom*. University of Chicago Press.

Gardner, David P., et al. 1983. *A Nation at Risk*. Report of the National Commission on Excellence in Education, US Department of Education, Washington, DC. ERIC, Institute of Education Sciences.

Greene, Jay, Marcus Winters, and Greg Forster. 2004. “Testing High-Stakes Tests: Can We Believe the Results of Accountability Tests?” *Teachers College Record* 106(6): 1124–44.

Hamilton, Laura, Mark Berends, and Brian M. Stecher. 2005. “Teachers’ Responses to Standards-Based Accountability.” RAND Working Paper WR-259-EDU.

Hanushek, Eric A., and Margaret F. Raymond. 2005. “Does School Accountability Lead to Improved Student Performance?” *Journal of Policy Analysis and Management* 24(2): 297–327.

Hastings, Justine S., and Jeffrey M. Weinstein. 2008. “Information, School Choice, and Academic Achievement: Evidence from Two Experiments.” *Quarterly Journal of Economics* 123(4): 1373–1414.

Heckman, James, Carolyn Heinrich, and

Jeffrey Smith. 1997. “Assessing the Performance of Performance Standards in Public Bureaucracies.” *American Economic Review* 87(2): 389–95.

Heckman, James J., Carolyn Heinrich, and Jeffrey Smith. 2002. “The Performance of Performance Standards.” *Journal of Human Resources* 37(4): 778–811.

Hershbein, Brad, and Melissa Kearney. 2014. “Major Decisions: What Graduates Earn Over their Lifetimes.” The Hamilton Project, Brookings Institution. http://www.hamiltonproject.org/papers/major_decisions_what_graduates_earn_over_their_lifetimes/.

Hoekstra, Mark. 2009. “The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach.” *Review of Economics and Statistics* 91(4): 717–24.

Holmstrom, Bengt, and Paul Milgrom. 1991. “Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design.” *Journal of Law, Economics, and Organization*, 7(Special issue: [Papers from the Conference on the New Science of Organization, January 1991]): 24–52.

Hout, Michael, and Stuart W. Elliott. 2011. *Incentives and Test-based Accountability in Education*. National Academies Press.

Hoxby, Caroline M. 2003. “School Choice and School Productivity (Or, Could School Choice Be a Rising Tide that Raises All Boats?)” Chap. 8 in *The Economics of School Choice*, edited by C. Hoxby. University of Chicago Press.

Hoxby, Caroline M. 2009. “The Changing Selectivity of American Colleges.” *Journal of Economic Perspectives* 23(4): 95–118.

Hussain, Iftikhar. 2015. “Subjective Performance Evaluation in the Public Sector: Evidence from School Inspections.” *Journal of Human Resources* 50(1): 189–221.

Jacob, Brian A. 2005. “Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools.” *Journal of Public Economics* 89(5–6): 761–96.

Jacob, Brian A., and Steven D. Levitt. 2003. “Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating.” *Quarterly Journal of Economics* 118(3): 843–77.

Jin, Ginger Zhe, and Phillip Leslie. 2003. “The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards.” *Quarterly Journal of Economics* 118(2): 409–51.

Kane, Thomas J., and Douglas O. Staiger. 2002. “The Promise and Pitfalls of Using Imprecise School Accountability Measures.” *Journal of Economic Perspectives* 16(4): 91–114.

Kurlaender, Michal, Scott Carrell, and Jacob Jackson. 2016. “The Promises and Pitfalls of

Measuring Community College Quality." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2(1): 174–90.

Ladd, Helen F. 1999. "The Dallas School Accountability and Incentive Program: An Evaluation of Its Impacts on Student Outcomes." *Economics of Education Review* 18(1): 1–16.

Ladd, Helen F., and Douglas L. Lauen. 2010. "Status versus Growth: The Distributional Effects of School Accountability Policies." *Journal of Policy Analysis and Management* 29(3): 426–50.

Lauen, Douglas Lee, and S. Michael Gaddis. 2012. "Shining a Light or Fumbling in the Dark? The Effects of NCLB's Subgroup-Specific Accountability on Student Achievement." *Educational Evaluation and Policy Analysis* 34(2): 185–208.

Lemons, Richard, Thomas F. Luschei, and Leslie Santee Siskin. 2003. "Leadership and the Demands for Standards-based Accountability." Chap. 4 in *The New Accountability: High Schools and High-Stakes Testing*, edited by Martin Carnoy, Richard Elmore, Leslie Santee Siskin, 99–128. RoutledgeFalmer.

Looney, Adam, and Constantine Yannelis. 2015. "A Crisis in Student Loans? How Changes in the Characteristics of Borrowers and in the Institutions They Attended Contributed to Rising Loan Defaults." Brookings Paper on Economic Activity, Fall 2015 Conference.

Lu, Susan Feng. 2012. "Multitasking, Information Disclosure, and Product Quality: Evidence from Nursing Homes." *Journal of Economics and Management Strategy* 21(3): 673–705.

Luca, Michael, and Jonathan Smith. 2013. "Salience in Quality Disclosure: Evidence from the U.S. News College Rankings." *Journal of Economics and Management Strategy* 22(1): 58–77.

Macartney, Hugh. 2016. "The Dynamic Effects of Educational Accountability." *Journal of Labor Economics* 34(1): 1–28.

MacLeod, W. Bentley. 2003. "Optimal Contracting with Subjective Evaluation." *American Economic Review* 93(1): 216–40.

MacLeod, W. Bentley, Evan Riehl, Juan E. Saavedra, and Miguel Urquiola. 2015. "The Big Sort: College Reputation and Labor Market Outcomes." NBER Working Paper 21230.

MacLeod, W. Bentley, and Miguel Urquiola. 2015. "Reputation and School Competition." *American Economic Review* 105(11): 3471–88.

Meredith, Marc. 2004. "Why Do Universities Compete in the Rankings Game? An Empirical Analysis of the Effects of the *US News and World Report* College Rankings." *Research in Higher Education* 45(5): 443–61.

Mintrop, Heinrich, and Tina Trujillo. 2005. "Corrective Action in Low Performing Schools:

Lessons for NCLB Implementation from First-Generation Accountability Systems." *Education Policy Analysis Archives* 13(48).

Monks, James, and Ronald G. Ehrenberg. 1999. "The Impact of *U.S. News and World Report* College Rankings on Admissions Outcomes and Pricing Policies at Selective Private Institutions." NBER Working Paper 7227.

Murnane, Richard J. 2013. "US High School Graduation Rates: Patterns and Explanations." *Journal of Economic Literature* 51(2): 370–422.

Neal, Derek. 2010. "Aiming for Efficiency Rather than Proficiency." *Journal of Economic Perspectives* 24(3): 119–31.

Neal, Derek. 2013. "The Consequences of Using One Assessment System to Pursue Two Objectives." NBER Working Paper 19214.

Neal, Derek, and Diane Whitmore Schanzenbach. 2010. "Left Behind by Design: Proficiency Counts and Test-Based Accountability." *Review of Economics and Statistics* 92(2): 263–83.

Obama, Barack. 2015. "Weekly Address: A New College Scorecard." September 12. <https://www.whitehouse.gov/the-press-office/2015/09/12/weekly-address-new-college-scorecard>.

Oreopoulos, Philip, and Kjell G. Salvanes. 2011. "Priceless: The Nonpecuniary Benefits of Schooling." *Journal of Economic Perspectives* 25(1): 159–84.

Özek, Umut. 2012. "One Day Too Late? Mobile Students in an Era of Accountability." Working Paper 82, CALDER.

Reback, Randall. 2008. "Teaching to the Rating: School Accountability and the Distribution of Student Achievement." *Journal of Public Economics* 92(5–6): 1394–1415.

Reback, Randall, Jonah Rockoff, and Heather L. Schwartz. 2014. "Under Pressure: Job Security, Resource Allocation, and Productivity in Schools under NCLB." *American Economic Journal: Economic Policy* 6(3): 207–41.

Rockoff, Jonah, and Lesley J. Turner. 2010. "Short-Run Impacts of Accountability on School Quality." *American Economic Journal: Economic Policy* 2(4): 119–147.

Rouse, Cecilia Elena, Jane Hannaway, Dan Goldhaber, and David Figlio. 2013. "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure." *American Economic Journal: Economic Policy* 5(2): 251–81.

Stecher, Brian M., Sheila Barron, Tammi Chun, and Karen E. Ross. 2000. *The Effects of the Washington State Education Reform on Schools and Classrooms*. Santa Monica, CA: RAND Corporation.

Toenjes, L. A., and Garst, J. E. 2000. "Identifying High Performing Texas Schools and School

Districts and their Methods of Success.” Texas Education Agency.

US Department of Education. 2015. “Fact Sheet: Obama Administration Increases Accountability for Low-Performing For-Profit Institutions.” July 1. <http://www.ed.gov/news/press-releases/fact-sheet-obama-administration-increases-accountability-low-performing-profit-institutions>.

White House. 2013. “Fact Sheet on the President’s Plan to Make College More Affordable: A Better Bargain for the Middle Class.” August 22. <https://www.whitehouse.gov/the-press-office/>

2013/08/22/fact-sheet-president-s-plan-make-college-more-affordable-better-bargain-.

Winston, Gordon C. 1999. “Subsidies, Hierarchy and Peers: The Awkward Economics of Higher Education.” *Journal of Economic Perspectives* 13(1): 13–36.

Winters, Marcus A., Julie R. Trivitt, and Jay P. Greene. 2010. “The Impact of High-Stakes Testing on Student Proficiency in Low-Stakes Subjects: Evidence from Florida’s Elementary Science Exam.” *Economics of Education Review* 29(1): 138–46.

