

- Polikoff, M., McEachin, A., Wrabel, S., & Duque, M. (2015). Grading the No Child Left Behind waivers. Washington, DC: American Enterprise Institute.
- Reback, R., Rockoff, J., & Schwartz, H. (2014). Under pressure: Job security, resource allocation and productivity in schools under No Child Left Behind. *American Economic Journal: Economic Policy*, 6, 207–241.
- Sutcher, L., Darling-Hammond, L., & Carver-Thomas, D. (2016). A Coming Crisis in Teaching? Teacher Supply, Demand, and Shortages in the U.S. Report of the Learning Policy Institute. Palo Alto, CA: Learning Policy Institute.
- Usher, A. (2015). AYP results for 2010-11 (Rep.). Retrieved, from files.eric.ed.gov/fulltext/ED527525.pdf. (accessed January 17, 2017)

THE CHANGING FEDERAL ROLE IN SCHOOL ACCOUNTABILITY

Brian Jacob

When President George W. Bush signed the *No Child Left Behind Act* (NCLB) in 2002, it marked a historic expansion of the federal government’s role in U.S. education policy. NCLB had a broad and deep impact on education policy and practice throughout the country.

One of the most immediate and visible effects of NCLB was the requirement that schools administer standardized exams in reading and math in grades 3 to 8. Prior to the passage of NCLB, testing was often determined at the district level, administered in only select grades, and not consistently used by school or district leaders. The legislation also required states to report student performance for each school annually, indicating the fraction of students meeting proficiency standards overall and separately for a variety of subgroups. Mandated subgroups included traditional race and gender categories, as well as categories for economically disadvantaged, limited English proficient, and special needs students.

The legislation required schools to increase the fraction of students meeting proficiency each year in order to attain the goal of 100 percent proficiency by 2014. Schools failing to meet these goals were designated as not making Adequate Yearly Progress (AYP), and subject to an increasingly severe set of sanctions, which ranged from the requirement to develop a school improvement plan to a complete restructuring of the school.¹

THE THEORETICAL FRAMEWORK OF SCHOOL ACCOUNTABILITY

The underlying rationale for school accountability policies such as NCLB stems from what economists refer to as a principal-agent problem. The idea is that

¹ NCLB included several other accountability provisions, including requirements to provide all students with a “highly qualified” teacher and to allow students in schools failing to meet AYP to obtain supplemental education services and exercise school choice. In practice, these provisions had little impact on schools and quickly disappeared from public discourse.

because the principal (i.e., stakeholders in the educational system such as parents and policymakers) has difficulty monitoring day-to-day activity within schools, it is possible for the agents (i.e., teachers or school personnel) to act in ways that do not advance the interest of the stakeholders. For example, teachers might spend more time on subjects such as art or social studies that policymakers view as less important than math and reading. Alternatively, principals may eschew the time-consuming and socially uncomfortable task of removing an ineffective teacher.

An accountability system will establish clear, measurable, and ambitious performance standards; seek to align curriculum and instruction to these standards; assess students to measure their progress in meeting the standards; and attach rewards or sanctions to school performance.² However, as many observers have noted, accountability policies also provide incentives for schools to engage in a host of behaviors that might interfere with student learning even while helping the school meet its performance targets. For example, school accountability systems provide educators an incentive to reduce time devoted to low-stakes subjects or grades (e.g., art, music, social studies), to shift effort toward students whose performance is most consequential for the school's rating (which is often students close to the proficiency threshold), teach narrowly to the content most likely to be assessed in the high-stakes exam, or reclassify low-performing students into programs where they are exempt from testing (e.g., special education).

The goal of this article is to discuss the influence of school accountability policies on educational inputs (e.g., funding, teacher quality, instructional practices, and school organization) as well as outputs (e.g., student achievement). I will first review the evidence on NCLB as well as earlier state accountability systems, discussing the evidence for both positive and negative effects and reviewing what we still have left to learn with regard to these policies. I will then summarize what we know about the potential effects of the current federal approach to school accountability. Finally, I offer some thoughts about what role the federal government should play in school accountability.

THE GOOD

School accountability has had several positive effects. First, the requirement that schools regularly test all students and report results separately by subgroup has illuminated large and persistent inequities in student performance, many of which had been ignored in the past. In my judgment, this is one of the most important impacts of the legislation. The proliferation of such data has not only galvanized advocacy groups and spurred policy reforms, but also fostered new research studying the predictors of student performance and the effectiveness of various interventions.³

Second, the existing research suggests that test-based accountability has had a modest positive effect on student performance. There is a vast body of research on this issue and a comprehensive review of the literature is beyond the scope of this article. I highlight several key findings below, and refer interested

² As Figlio and Loeb (2011) recognize, accountability systems can be effective even if the threatened sanctions are not credible. The public disclosure of student performance information alone can induce considerable pressure for reform (see, e.g., Figlio & Lucas, 2004).

³ Unfortunately, NCLB did not require states to use the same assessment and allowed states considerable latitude in defining proficiency. As a result, states adopted tests of varying quality and set wildly different standards for what constituted proficiency.

readers to the review pieces by Figlio and Ladd (2015) and Figlio and Loeb (2011).

One large set of studies examines the effect of accountability threats and sanctions on student achievement by comparing student performance (or changes in performance) in sanctioned versus (otherwise similar) non-sanctioned schools. For example, Chiang (2009) uses a regression discontinuity design to study the impact of receiving a failing grade in Florida's pre-NCLB accountability system.⁴ He finds modest positive effects that are larger in math, on the high-stakes exam and among previously low-performing students. Perhaps most interestingly, he finds that the threat-induced improvements to math scores in elementary school persist into middle school. More recent studies have taken the same approach to studying NCLB, and typically find modest positive effects (see, e.g., Ahn & Vigdor, 2014; Chakrabarti, 2014; Reback, Rockoff, & Schwartz, 2014).

Because these studies compare schools on the margin of being sanctioned, they do not capture the full systemic effects of school accountability. Several studies attempt to measure these broader impacts of accountability by examining the trends in student performance across states over time (Carnoy & Loeb, 2002; Dee & Jacob, 2011; Hanushek & Raymond, 2005). An additional advantage of these studies is that they rely on the National Assessment of Educational Progress (NAEP)—a low-stakes exam with broad content coverage and good psychometric properties—to measure student performance. The consensus from this research is that school accountability led to increases in elementary math scores. For example, Dee and Jacob (2011) study the impact of NCLB using a comparative interrupted time series analysis that relies on comparisons of the test-score changes across states that already had school accountability policies in place prior to NCLB and those that did not. They find positive effects in math on average, as well as at the top and bottom of the ability distribution, and no evidence of a negative effect on science performance (a low-stakes subject).⁵

It should be noted, however, that the effects vary considerably based on the context, nature of the sanctions, and so on. For example, several studies find that student performance only increases in the subject for which the school was sanctioned, and there is evidence that intermediate sanctions have less impact than either initial sanctions or the final, most severe sanctions.

Second, there is evidence that the early pressure introduced by school accountability led to some positive changes in terms of school organization and instructional practices. For example, using data from principal surveys administered over time, Rouse et al. (2013) examine how pressure under Florida's pre-NCLB accountability system influenced a host of school policies and practices. They find that schools receiving an "F" rating made a variety of substantial changes, such as focusing more attention on at-risk students through mandatory before- and after-school tutoring, lengthening the time devoted to instruction, increasing common planning and individual prep time for teachers, increasing resources to teachers, and making organizational changes such as introducing block scheduling.⁶ In contrast to many anecdotal accounts, Grissom, Nicholson-Crotty, and Harrington (2014) find that teacher reports of various work environment measures improved over the period coinciding with the implementation of NCLB. When they compare trends in states with and without pre-NCLB accountability systems, they find little evidence that

⁴ A number of other studies have found comparable results when studying the Florida system (see Figlio & Loeb, 2011).

⁵ Non-linear trends in reading performance over the 1990s prohibit the authors from confidently measuring the effects of NCLB on reading performance.

⁶ See also Reback et al. (2014).

these measures changed in response to NCLB. At the same time, the studies that report positive effects focus on the years immediately following the introduction of accountability systems, and there was considerable and growing dissatisfaction with NCLB as the required performance levels rose.

THE BAD

As with most educational policies, school accountability has had some negative consequences. There is substantial evidence that schools strategically responded to accountability in ways to maximize a school's rating without moving underlying student learning. Researchers have documented that under pre-NCLB accountability systems, schools (a) reclassified students into programs such as special education where they would not be subject to accountability provisions (Jacob, 2005), (b) fed students high-calorie food to boost performance on test day (Figlio & Winicki, 2005), (c) strategically disciplined low-performing students so that they were not in school during the testing window (Figlio, 2006), and (d) changed student answers on high-stakes exams (Jacob & Levitt, 2003).

While some of the more egregious practices were less common under NCLB because of stricter test security along with the requirement that schools test at least 95 percent of students, there is evidence that schools focused efforts on high-stakes grades and subjects, and began "teaching to the test"—that is, narrowly tailoring instruction to focus on content most likely to appear on the high-stakes exam. As a result, student performance on high-stakes exams is typically much higher than on low-stakes measures of the same subject (see, e.g., Jacob, 2005; Jennings & Bearak, 2014; Jennings & Lauen, 2016).⁷ At the very least, this represents an inefficient allocation of teacher effort and school resources; at worst, one might worry that all of the gains were ephemeral.

Interestingly, while there is good evidence that these policies have induced schools to shift time away from low-stakes subjects such as science and social studies (Dee, Jacob, & Swartz, 2013; Reback, Rockoff, & Schwartz, 2014), it does not appear that there have been any negative impacts on student test performance in these areas (see, e.g., Dee & Jacob, 2011; Winters, Trivitt, & Greene, 2010). It is possible that the improvements in reading and math had positive spillovers into these subjects. On the other hand, it is hard to quantify the impact of less time on subjects such as art, music, and physical education.

THE UNKNOWN

A critical question with respect to school accountability, and NCLB in particular, is how it influenced the *distribution* of student achievement. That is, how did it help low-performing or economically disadvantaged students or students of color relative to other students? Many observers have noted that accountability systems such as NCLB provide schools an incentive to focus attention on students close to the proficiency threshold, and potentially neglect particularly low- or high-performing students.

While dozens of studies have examined this issue of "educational triage," the results are all over the board. Several studies find evidence consistent with such

⁷ Fuller and Ladd (2013) find evidence that schools in North Carolina strategically reassigned more qualified teachers from lower (i.e., untested) to higher (i.e., tested) elementary grades following the introduction of school accountability policies.

targeting (Krieg, 2008; Neal & Schanzenbach, 2010; Reback, 2008), but many other studies do not (Ballou & Springer, forthcoming; Chakrabarti, 2014; Dee & Jacob, 2011; Gaddis & Lauen, 2014; Ladd & Lauen, 2010; Reback, Rockoff, & Schwartz, 2014; Springer, 2008).

One potential reason for the inconsistent empirical findings is that the marginal student is hard to identify in practice, and likely depends on a variety of factors including the level of the proficiency threshold (Ballou & Springer, forthcoming; Lauen & Gaddis, 2016). In addition, most studies do not measure school effort directly, but rather focus on student performance across different groups. To the extent that some school actions have spillovers to inframarginal students, or students are differentially responsive to similar inputs, these results can be difficult to interpret. Finally, studies that use low-stakes exams typically do not find such evidence, suggesting that the additional attention paid to students near the cutoff comes in the form of test-specific coaching (Jennings & Sohn, 2014).

So, while certain accountability systems at certain times may have differentially benefited a narrow set of students, the weight of the evidence is that accountability systems have produced gains for most students, with the largest gains accruing to low-performing or traditionally disadvantaged groups.

Another important question is how school accountability policies influence student outcomes in the long run. Only one large, quantitative analysis addresses this issue. Deming et al. (forthcoming) study the long-run impact of Texas' school accountability system in the 1990s that served as the model for NCLB. They find that students who attended high schools in years when the school was subject to greater accountability pressure not only had higher 10th-grade test scores, but also were more likely to have attended college, completed a four-year degree, and had higher earnings at age 25. These long-term benefits were driven by students with low 8th-grade test scores. Unfortunately, there is no evidence on the long-term impacts of accountability in elementary or middle schools.

Finally, as mentioned above, there is not much systematic evidence on how school accountability influenced school-level policies and practices several years after the initial introduction of the policies. And there is little evidence on how accountability influenced high school organization or outcomes more generally.

THE FUTURE

In 2015, Congress passed the *Every Student Succeeds Act* (ESSA), which replaced NCLB's system of school accountability with a more narrowly focused and less highly prescribed approach. ESSA maintains the testing and reporting provisions established under NCLB. But instead of requiring all schools to meet annual performance targets, ESSA requires states to focus on a small set of persistently low-performing schools and allows states and districts greater flexibility to design the interventions they deem appropriate.

While it is too early to assess the impact of ESSA itself, there have been a number of earlier efforts to turn around low-performing schools, including the School Improvement Grants provided under Race to the Top and the reforms mandated as part of NCLB waivers granted to virtually all states in 2012. Unfortunately, the evidence thus far suggests that these reforms have not been successful in raising student achievement.

A recent report from the U.S. Government Accountability Office found that many states did not effectively implement the waiver-based reforms, and had difficulty monitoring districts and schools (Nowicki, 2016). Recent evaluations of Priority and Focus school reforms in several different states found that the efforts had no impact on student performance levels or achievement gaps (Bonilla & Dee, 2016;

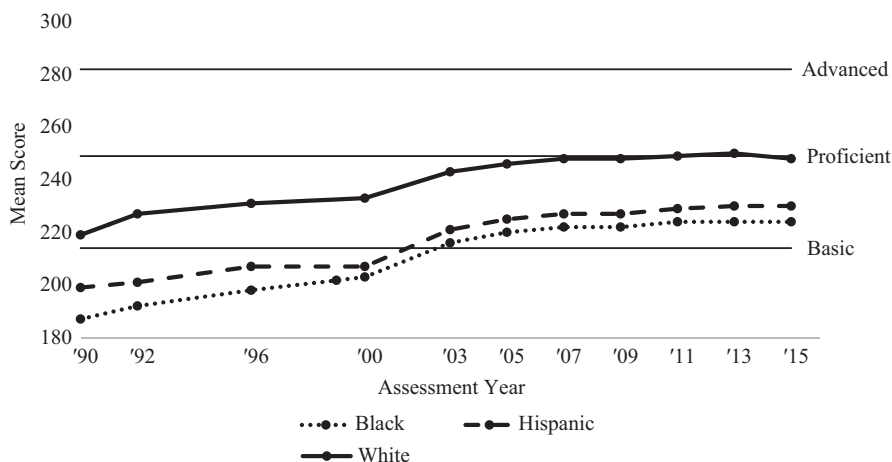


Figure 1. Trends in 4th-Grade Math NAEP Scores, by Race/Ethnicity.

Dee & Dizon-Ross, 2016; Dougherty & Weiner, 2016; Hemelt & Jacob, 2016). And the broader research on school turnaround efforts is not particularly encouraging either. While there have been a few successes (Ruble, 2015; Schueler, Goodman, & Deming, 2016), in many instances state and district efforts have failed to produce any notable improvement in student performance (Heissel & Ladd, 2016; Strunk et al., 2016; Zimmer, Henry, & Kho, 2016).

While it is possible that the modest additional funding available for reform efforts under ESSA may lead to better outcomes, the flexibility afforded in the new system likely means that states and districts with the least capacity—exactly where many of the most disadvantaged students reside—will have the most difficulty implementing the ambitious reforms needed to impact achievement.

CONCLUSIONS

School accountability has played a prominent role in federal and state education policy for at least 20 years. In my view, these policies have had modest positive effects on student achievement, particularly on math scores in typically low-performing schools. At the same time, it is clear that schools have responded strategically to accountability policies. In some cases, the responses were undoubtedly counterproductive (e.g., cheating, exclusion of students, the most narrow forms of test preparation); in other cases, the responses may simply reflect different priorities (e.g., shift toward math and reading, strategic reassignment of teachers).

Figure 1 illustrates the trends in 4th-grade math scores for white, black, and Hispanic students in the United States since 1992. Performance increased among all groups through the mid-2000s, at which point it started to level off, and then dropped in 2015.⁸ The same pattern is evident at different percentiles and is roughly comparable for eighth graders. Improvement in reading has not been as rapid as in math, though the pattern is similar (see Fig. 2).

These trends indicate that policy has a long way to go toward ensuring that all students are prepared for the 21st-century job market. I would argue that some type

⁸ For more information on the levels and trends of student performance, see the NAEP results at https://www.nationsreportcard.gov/reading_math_2015/#?grade=4.

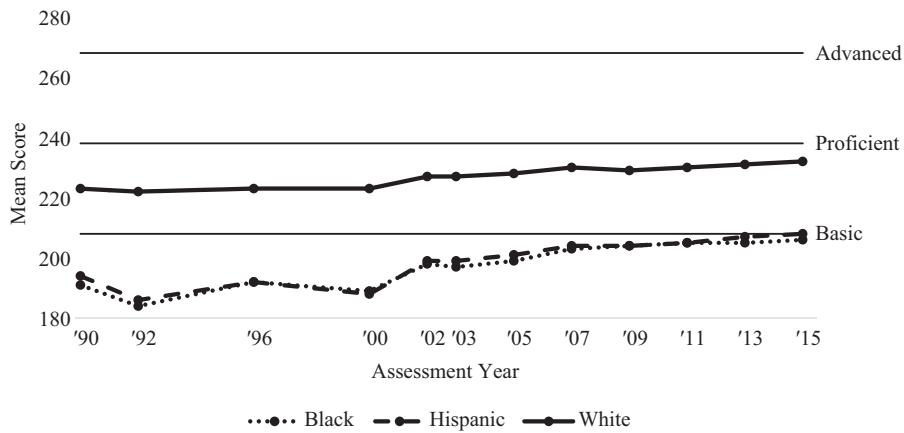


Figure 2. Trends in 4th-Grade Reading NAEP Scores, by Race/Ethnicity.

of school accountability that measures and publicly reports student performance separately for specific populations is a necessary though not sufficient condition for improvement. I am certain that teachers, administrators, and legislators will be happier with the more limited and flexible approach to accountability embodied in ESSA. I am skeptical that this approach will lead to meaningful improvement in student performance. I fear that the more affluent districts and states, those with more human capital or smoothly functioning political structures, will fare well, while others will be left behind.

BRIAN JACOB is the Walter H. Annenberg Professor of Education Policy and a Professor of Economics in the Gerald R. Ford School of Public Policy at the University of Michigan, 735 South State Street, Ann Arbor, MI 48109 (e-mail: bajacob@umich.edu). He is also a Research Associate for the National Bureau of Economic Research.

REFERENCES

- Ahn, T., & Vigdor, J. (2014). The impact of No Child Left Behind's accountability sanctions on school performance: Regression discontinuity evidence from North Carolina. Working paper no. 20511. National Bureau of Economic Research. Cambridge, MA.
- Ballou, D., & Springer, M. G. (2017). Has NCLB encouraged educational triage? Accountability and the distribution of achievement gains. *Education Finance and Policy*, forthcoming.
- Bonilla, S., & Dee, T. S. (2016). The impact of NCLB waiver focus school reforms on achievement gaps: Evidence from Kentucky. Working Paper. Stanford University.
- Burnette II, D. (2016). States, districts to call shots on turnarounds under ESSA. *Education Week*. Retrieved October 16, 2016, from <http://www.edweek.org/ew/articles/2016/01/06/states-districts-to-call-shots-on-turnarounds.html>.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24, 305–331.
- Chakrabarti, R. (2014). Incentives and responses under No Child Left Behind: Credible threats and the role of competition. *Journal of Public Economics*, 110, 124–146.
- Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics*, 93, 1045–1057.
- Dee, T. S., & Dizon-Ross, E. (2016). School performance, accountability and waiver reforms: Evidence from Louisiana. Working paper. Stanford university.

- Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30, 418–446.
- Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, 35, 252–279.
- Deming, D. J., Cohodes, S., Jennings, J., & Jencks, C. School accountability, postsecondary attainment and earnings. *Review of Economics and Statistics*, forthcoming.
- Dougherty, S. M., & Weiner, J. (2016). The Rhode to turnaround? The impact of waivers to No Child Left Behind on school performance. Working paper. University of Connecticut.
- Figlio, D. N. (2006). Testing, crime and punishment. *Journal of Public Economics*, 90, 837–851.
- Figlio, D., & Ladd, H. F. (2015). School accountability and student achievement. In H. F. Ladd & M. Goertz (Eds.), *Handbook of research in education finance and policy* (2nd ed.) (194–210). New York: Routledge.
- Figlio, D., & Loeb, S. (2011). School accountability. In E. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 3, pp. 383–421). North Holland. Retrieved from https://cepa.stanford.edu/sites/default/files/Accountability_Handbook.pdf.
- Figlio, D. N., & Lucas, M. E. (2004). What's in a grade? School report cards and the housing market. *American Economic Review*, 94, 591–604.
- Figlio, D. N., & Winicki, J. (2005). Food for thought: The effects of school accountability plans on school nutrition. *Journal of Public Economics*, 89, 381–394.
- Fuller, S. C., & Ladd, H. F. (2013). School-based accountability and the distribution of teacher quality across grades in elementary school. *Education Finance and Policy*, 8, 528–559.
- Gaddis, S. M., & Lauen, D. L. (2014). School accountability and the black–white test score gap. *Social Science Research*, 44, 15–31.
- Grissom, J. A., Nicholson-Crotty, S., & Harrington, J. R. (2014). Estimating the effects of No Child Left Behind on teachers' work environments and job attitudes. *Educational Evaluation and Policy Analysis*, 36, 417–436.
- Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24, 297–327.
- Heissel, J. A., & Ladd, H. F. (2016). School turnaround in North Carolina: A regression discontinuity analysis. CALDER working paper. Retrieved 1/6/2017, from <http://www.caldercenter.org/sites/default/files/WP%20156.pdf>.
- Hemelt, S., & Jacob, B. (2016). Differentiated accountability and education production: Evidence from NCLB waivers in Michigan. Working paper. University of Michigan.
- Jacob, B. (2005). Accountability, incentives and behavior: Evidence from school reform in Chicago. *Journal of Public Economics*, 89, 761–796.
- Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics*, 118, 843–877.
- Jennings, J. L., & Bearak, J. M. (2014). “Teaching to the test” in the NCLB era: How test predictability affects our understanding of student performance. *Educational Researcher*, 43, 381–389.
- Jennings, J. L., & Lauen, D. L. (2016). Accountability, inequality, and achievement: The effects of the No Child Left Behind act on multiple measures of student learning. *Russell Sage Foundation Journal*, 2, 220–241.
- Jennings, J., & Sohn, H. (2014). Measure for measure how proficiency-based accountability systems affect inequality in academic achievement. *Sociology of Education*, 87, 125–141.
- Krieg, J. M. (2008). Are students left behind? The distributional effects of the No Child Left Behind act. *Education Finance and Policy*, 3, 250–281.
- Ladd, H. F., & Lauen, D. L. (2010). Status versus growth: The distributional effects of school accountability policies. *Journal of Policy Analysis and Management*, 29, 426–450.
- Lauen, D. L., & Gaddis, S. M. (2016). Accountability pressure, academic standards, and educational triage. *Educational Evaluation and Policy Analysis*, 38, 127–147.

- Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics*, 92, 263–283.
- Nowicki, J. M. (2016). K-12 education: Education's experiences with flexibility waivers could inform efforts to assist states with new requirements (No. GAO-16-650). Washington, DC: United States Government Accountability Office. Retrieved 1/6/2017, from <http://www.gao.gov/products/GAO-16-650>.
- Reback, R. (2008). Teaching to the rating: School accountability and the distribution of student achievement. *Journal of Public Economics*, 92, 1394–1415.
- Reback, R., Rockoff, J., & Schwartz, H. L. (2014). Under pressure: Job security, resource allocation, and productivity in schools under No Child Left Behind. *American Economic Journal: Economic Policy*, 6, 207–241.
- Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2013). Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure. *American Economic Journal: Economic Policy*, 5, 251–281.
- Ruble, W. (2015). The effect of contracting out low performing schools on student performance. Working paper no. 1521. Tulane University, Department of Economics. Retrieved from <https://ideas.repec.org/p/tul/wpaper/1521.html>.
- Schueler, B. E., Goodman, J., & Deming, D. J. (2016). Can states take over and turn around school districts? Evidence from Lawrence, Massachusetts. Working paper no. 21895. National Bureau of Economic Research. Cambridge, MA.
- Strunk, K. O., Marsh, J. A., Hashim, A. K., Bush-Mecenas, S., & Weinstein, T. (2016). The impact of turnaround reform on student outcomes: Evidence and insights from the Los Angeles unified school district. *Education Finance and Policy*, 11, 251–282.
- Springer, M. G. (2008). The influence of an NCLB accountability plan on the distribution of student test score gains. *Economics of Education Review*, 27, 556–563.
- Winters, M. A., Trivitt, J. R., & Greene, J. P. (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, 138–146.
- Zimmer, R., Henry, G. T., & Kho, A. (2016, May). The role of governance and management in school turnaround policies: The case of Tennessee's achievement school district and iZones. Paper presented at the Association of Education Finance and Policy, Denver, CO. Retrieved 1/6/2017, from https://aefpweb.org/sites/default/files/webform/41/ASD_iZone%20Impact%20Paper%20conference%20online%20version.pdf.

NCLB: RESPONSE TO JACOB

Helen F. Ladd

Brian Jacob has done an excellent job of summarizing the research literature on school accountability programs in the United States. Importantly, though, much of the evidence he cites relates to state programs that were initiated prior to the passage of No Child Left Behind (NCLB) at the national level. These state programs often were part of more comprehensive reform packages and took different forms than the federal NCLB law. For example, many of the state accountability systems were based on average levels or gains in test scores rather than on the proficiency rates