

Answering the Call?

EXPLAINING HOW STATES HAVE (OR HAVE NOT) TAKEN UP
THE ESSA ACCOUNTABILITY CHALLENGE

PAUL MANNA AND ARNOLD SHOBER

ESSA's Theory of Action

Education-policy aficionados can count on two things accompanying each revision to the Elementary and Secondary Education Act (ESEA). One is brimming optimism from supporters of the revised law, who assure the nation and potential skeptics that this time, yes, *this time* ESEA will deliver on its promise to advance educational equity and excellence. The 2015 reauthorization, dubbed the Every Student Succeeds Act (ESSA), was no exception. At its signing, President Obama proudly declared that ESSA would “make sure that every student is prepared to succeed in the 21st century,” a sentiment that supporters in Congress and the states echoed.¹

In general, when members of Congress reauthorize ESEA, they essentially recalibrate the federal role in education by dialing up or down Washington's involvement in American schools. ESSA's authors followed that playbook in attempting to correct the perceived overreach of No Child Left Behind (NCLB), the previous version of ESEA, which was many years overdue for revision.² In trying to craft a federal presence that was neither too hot nor too cold, lawmakers writing ESSA embraced a theory of action built around two main ideas.³

First, they assumed NCLB had laudable goals, but also had employed a heavy-handed approach that should be pulled back so states could exercise broader authority when implementing ESSA. The federal government's admirable ambitions led it to overreach with NCLB, imposing new requirements and procedures on states and local school districts. As a result, the argument went, NCLB forced subnational governments to follow Washington's lead rather than apply their own insights, experiences, and assessment of educational needs and aspirations to guide K–12 policy. A better approach, which ESSA attempted to embody, would scale back federal prescriptions for school accountability but not abandon them altogether.

Second, ESSA's authors assumed that state governments were capable innovators and, with the regulatory relief ESSA provided, would faithfully act to ensure all students



succeeded. A major line of state advocacy during the NCLB era, which encouraged the Obama administration to experiment aggressively with NCLB waivers, was that states needed greater flexibility from federal constraints to promote educational excellence and equity for all, goals that several states had themselves set in the 1980s.⁴ The authors of ESSA agreed. As a result, “accountability” remained a controlling idea in the new law, and it kept some of NCLB’s elements, while on others it permitted states to shift in new directions.

Typically, accountability systems embrace two broad categories of activities. The first category focuses on standards and measurement and involves establishing what students should know and be able to do (content standards), designing strategies for gauging performance (often involving testing), and setting up markers against which performance will be judged (performance standards). The second category focuses on which consequences should follow as a result of that measured performance. These consequences might include some combination of sanctions, supports, and rewards. How does ESSA compare with NCLB in these two broad categories?

In general, ESSA looks generally similar to NCLB in the first category and diverges sharply, but not completely, in the second. Like NCLB, the new law still requires states to test students annually in grades three to eight and to have a performance indicator for high school students. It also requires states to report student performance across various student subgroups (e.g., race, ethnicity, economic disadvantage, English learners, and others) at the state, district, and school level where there are sufficient numbers of students in each category. As with NCLB, states determine the N-size that would trigger whether a group’s results are reported. One shift with ESSA is that states are allowed to use nationally known standardized assessments, such as the SAT or ACT, instead of state assessments of high school reading, math, and science. Overall, though, there remained much continuity in the process of standard setting, testing, and measuring performance across NCLB and ESSA. For example, ESSA continued to provide states with flexibility to determine the content of their standards and tests, and states were enabled to make important administrative choices about how they presented results to the public and parents, a topic that we later describe in more detail.

The biggest shifts in accountability came in the second category of activities, in particular the consequences that flow from test results and other performance measures. The most striking difference was that ESSA significantly lessened the cascade of NCLB remedies that states were supposed to enforce on schools and districts

that failed to make adequate yearly progress (AYP). The law no longer requires that 100 percent of students meet proficiency goals by a certain date. Also gone are NCLB's requirements that schools missing AYP must offer their students public school choice and supplemental educational services (e.g., tutoring), and the additional corrective actions and potentially major restructuring to improve school performance that NCLB had demanded for schools that struggled during several consecutive years. ESSA's new approach focused on a smaller number of schools—dubbed schools requiring Comprehensive Support and Improvement (CSI), Targeted Support and Improvement (TSI), and Additional Targeted Support and Improvement (ATSI)—and gave states greater flexibility in designing various sanctions or supports to address those schools' needs. Although these new categories came with demands of their own, the states had their primary wish granted: namely, greater room to maneuver as they determined how to identify and react to schools that struggle to meet performance goals. Still, the aspects of NCLB that required transparency in reporting school performance results did remain, a form of accountability that makes information available so interested parties can press for explanations and potential changes when results decline. Some such requirements became even stiffer with ESSA, as with its requirement for reporting school-level financial data, something that previously had been essentially impossible to determine for anyone other than the most skilled professionals in the school finance field.⁵

State ESSA Plans as an Initial Test of the Law's Theory of Action

Once ESSA became law, a key question loomed: Would states leverage their new flexibility to develop innovative and effective approaches to create outstanding schools where all students achieved at high levels? State officials publicly promised to show that they were indeed committed to accountability while sacrificing neither excellence nor equity. Chris Minnich, then executive director of the Council of Chief State School Officers, argued that “our members want to be held accountable” and that state education agency leaders were “ready to build better accountability systems.”⁶ In a letter to the US Department of Education (ED) during ESSA's early implementation, the National Governors Association offered similar reassurances that “states and local school districts have made equity a priority and will use implementation of this new law to elevate that work,” while emphasizing the law's theory that “equity and flexibility can co-exist when states and school districts are afforded the ability to determine their own path to a high-quality education for all students.”⁷ Kristin Amundson, president and CEO of the National Association of State Boards of Education, echoed those sentiments in praising federal officials for their “willingness to work with states,” and she also promised that “states are strongly committed to



developing and implementing accountability systems that will ensure the success of every child.”⁸

State leaders and their representatives saying reassuring things about ESSA implementation is one thing. More difficult, though, is actually using the law’s new flexibility to improve schools. Historically, states have a mixed record, with some of their actions proving ineffective, and in some cases even harmful, in advancing educational excellence and equity. Such past failings are what led to Uncle Sam applying a heavier hand in schools in the first place. In sizing up ESSA, for example, Cynthia Brown, senior fellow at the Center for American Progress, argued that the law “virtually enshrines states’ rights, a governance structure with negative history in guaranteeing equal educational opportunities.”⁹ Why would ESSA’s lighter federal touch now produce higher levels of performance for all?

Although it is obviously premature to gauge ESSA’s effects on student achievement and school performance, one can begin to assess which states have adopted strategies that appear to keep their promise to hold fast to results-based accountability for schools. To that end, in this descriptive paper, we tackle two empirical questions about ESSA implementation and the forces shaping it. First, how committed to accountability are state ESSA plans? Simply describing what states have done is a first step toward understanding how policies have shifted in this new era. Second, which state-level factors are associated with a commitment to accountability? Knowing whether states are pursuing relatively tough approaches to accountability says nothing about why they have adopted such approaches. To begin unpacking these reasons, we examine potential forces that previous literature has shown can influence policy implementation. We draw on multiple sources of evidence and employ several methods to better understand what accountability looks like in state ESSA plans and which forces might be pushing some states to set accountability bars higher than others have done.

Our descriptive approach considers ESSA from three perspectives. First, we report results from fifty-state quantitative summaries of the accountability components of state ESSA plans, compiled by organizations active in the education policy space. Second, we examine four sets of factors that might explain these components: prior student achievement, prior policy commitments, features of state contexts likely to influence policy implementation, and state political contexts. These results provide basic bivariate relationships between the content of ESSA plans and factors that might account for it. Third, we augment our fifty-state analysis with brief case studies of

California, Florida, Iowa, Michigan, and New Mexico. The experiences of these five states provide a more granular look at the details of state plans and the forces that might have shaped them. Overall, we make no claims about causal relationships, but we suggest that our approach could help guide future studies interested in making causal arguments about the forces influencing the level of accountability that state ESSA plans demand.

ESSA Plan Content and Accountability

Most states' ESSA plans were approved relatively quickly by ED. Yet an overall thumbs-up or thumbs-down judgment from federal officials is a crude measure of quality. As noted earlier, given the intense interest in ESSA from scholars and policy practitioners, multiple organizations have evaluated the states' work, including their focus on accountability. We rely on three such sources in particular.

Bellwether Education Partners (Bellwether) is our first source. In its analysis, Bellwether invited forty-five policymakers and observers to code each state's ESSA plan on nine components.¹⁰ These included the quality of the state's proposed standards and assessment, the alignment of measures of success to the state's goals, and whether the state's plan provided incentives for schools to promote both student academic growth and academic proficiency. The second source is Alliance for Excellent Education (A4ED), which provided a stronger focus on accountability measures and specific components of state goals, largely related to achievement and graduation. A4ED's "Equity Dashboard" made its advocacy purposes clear. Its president, Bob Wise, noted that its work would help "teachers, parents, and advocates easily [to] identify whether states are . . . keeping students in the slow lane."¹¹ A third source, the Thomas B. Fordham Institute (Fordham), developed the most focused rating scale of state ESSA plans, emphasizing three components: clarity, applicability to all students (not just low performers), and awareness of the effects of poverty on achievement.¹²

Clearly, each organization had its own goals for ESSA and its own ideals for state implementation. In general, Bellwether, A4ED, and Fordham all supported standards-based public accountability. Furthermore, all can be seen as representing a DC-centric view of federal education policy rather than a perspective more deeply embedded in the states themselves. It could be that state policymakers, dealing as they must with actual local schools, districts, and board members, have a better handle on what "state flexibility" should mean. To foreshadow some of the discussion we surface in our case studies, perhaps teachers, boards, and even voters in California see greater



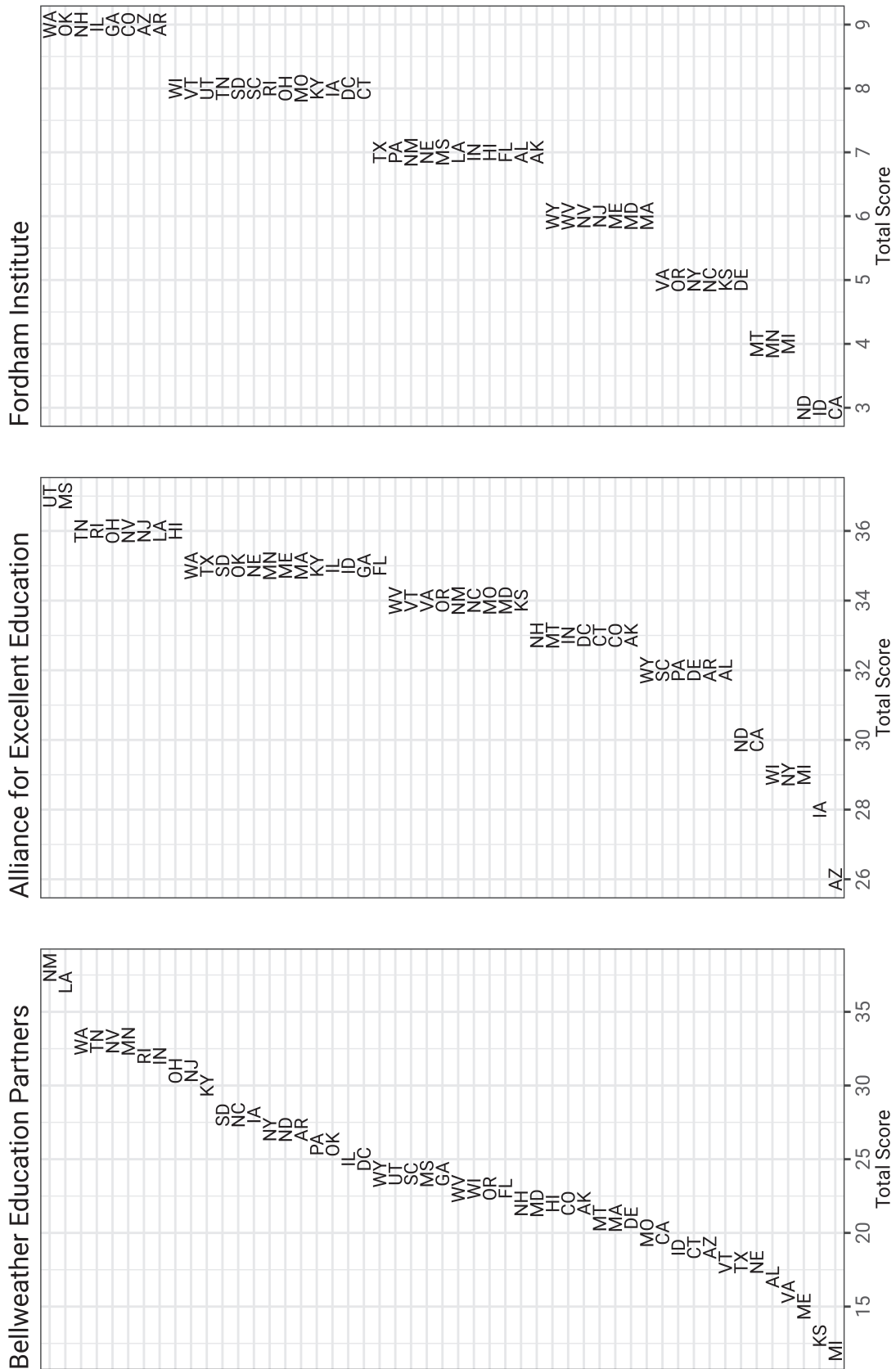
benefits from reporting multiple measures of education than they see in simple single-metric indices. Or perhaps New Mexico's Public Education Department has the pulse of teachers seeking flexibility and freedom in their work more than either ED or any DC-based education organization does. If ESSA was meant to send the "national school board" packing, in the words of Senator Lamar Alexander, perhaps these organizations have stacked the deck against the states, and against the purported goals of the law.

We argue that these metrics are still valid and reliable indices of what the states actually did. First, these organizations, and Bellwether in particular, constructed their ratings with a close eye on the actual content of state plans for implementation. They then did their best to reach judgments about those plans using a consistent set of criteria. For example, when Bellwether raters evaluated whether a state had "planned interventions . . . sufficiently rigorous to match the challenges those schools face," raters reviewed the specific items the state noted: what technical assistance the state would provide, who would provide the guidance, and which groups of students the state emphasized. That is, even if state-based advocates disagreed with Bellwether's rating, the rating still would reflect what state plans contained rather than functioning only as an interpretation of what was in the plan. State advocates might see a 2 from Bellwether as positive rather than negative, but the 2 still captures the *content* of ESSA plans in a consistent way.¹³

Second, we recognize that we measured the plans themselves rather than what the states actually implemented. The reason for this is that state plans represent the states' *intentions*. We were interested in what states would claim to do as a measure of how they took up the ESSA challenge in plan development, an early stage of implementation, before they ran into the buzz saw of on-the-ground realities. A state with limited goals and modest expectations might not be a source for best practices or ideas about new approaches, whereas a state with great expectations might invite further investigation to understand its successes or failures. Further, we are not measuring how well states have implemented their plans, or the effect of plans on students or schools. Those valuable questions are outside the scope of this study.

We begin considering the Bellwether, A4ED, and Fordham ratings in figure 1, which presents the relative rankings of state ESSA plans according to each organization's coding. The highest-ranked states appear atop each panel. The diversity of coders' focus is quite apparent, even though all rating systems are positively, albeit somewhat weakly, correlated.¹⁴

Figure 1. Ratings of state ESSA plans

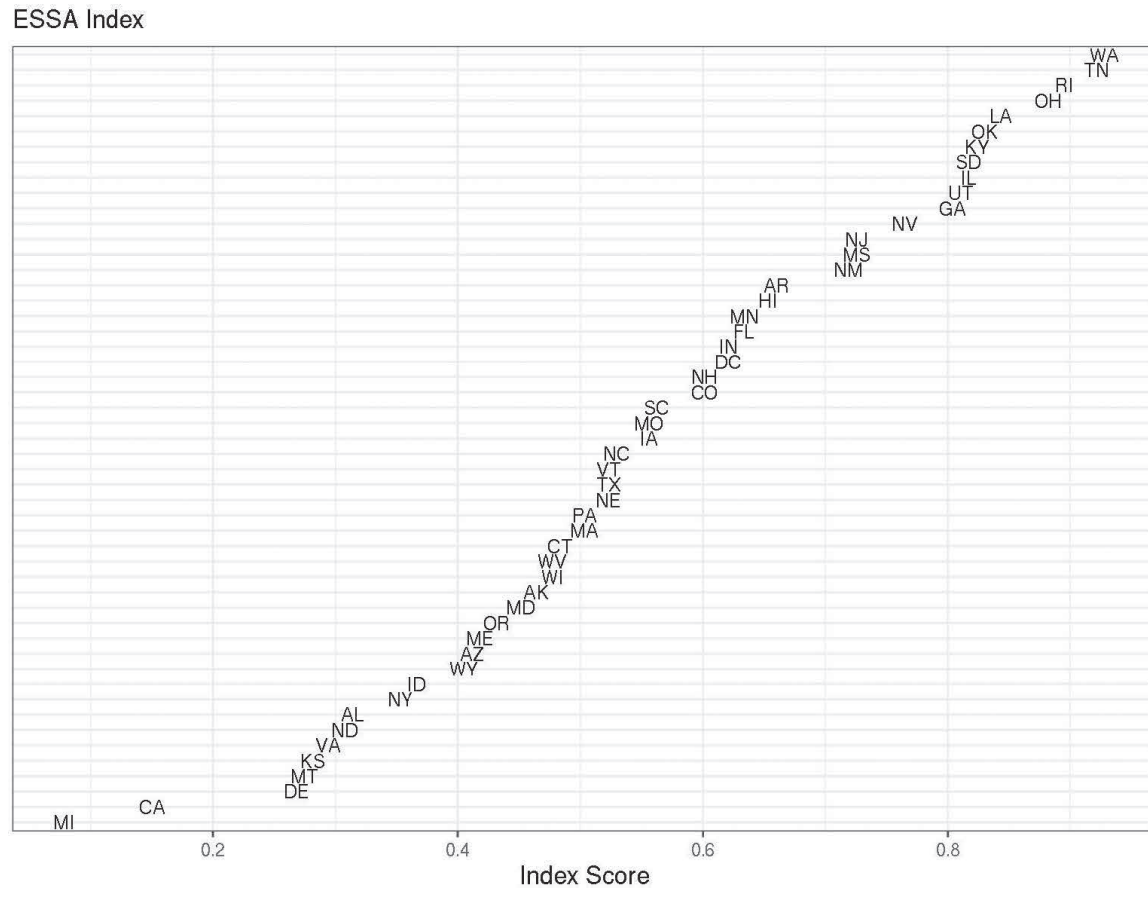


First, few states received truly low ratings from any organization, but low-scoring states were consistently low. Michigan, California, and Virginia received low ratings from all three organizations, for example. Second, states that focused on stakeholders and public accountability typically did quite well on both those dimensions (as is apparent in ratings from A4ED and Fordham). That is, states concentrated their plans where political pressure would be most likely to come from stakeholder groups, such as educator professional associations, teacher unions, and other groups interested in K–12 schooling. Third, most states took a path of little change, and ESSA implementation in those states appeared to be innovative neither in curriculum standards nor in choice of school quality measures. This is most apparent in Bellwether’s ratings, where many states appear with middling scores below the overall mean, but the pattern also appears from the other raters too.

We further analyzed ESSA plan ratings in figure 2, where we present a composite index developed from the scoring of Bellwether, A4ED, and Fordham. Individually, each captures a slightly different look at the states’ responses to ESSA, and combining them allows us to capture the variation in those responses. In essence, this index suggests an approximate net “quality” rating for each state’s ESSA plan.¹⁵ We detect two overall results in figure 2.

First, we note three rough groups in the data, which emerge where we see optical breaks among major clusters. Admittedly, the precise borders among these groups are debatable, so we suggest these groupings only as a rough heuristic. The first, at the high end, spans right to left from Washington to New Mexico. That grouping includes several states that have been at the center of debates over education reform, curriculum, and accountability. Louisiana, Ohio, Kentucky, and Oklahoma have attracted many reform and counter-reform movements over the past decade. They have not all taken the same steps, and they have not always moved toward common ends, but education policy has been a central political issue in all these states. At the lower end of the top cluster we see New Mexico, Mississippi, New Jersey, and Nevada. Three of these are home to traditionally disadvantaged populations and have taken strong steps to change course. We suggest that the second major cluster, from Arkansas to Wyoming, is a sort of default group. We note that the slope in the figure through this patch of figure 2 is fairly steep, indicating little net difference in ESSA plan scores from the lowest- to the highest-scoring state. A final group, at the low end of the spectrum (Idaho and below), completes the data. Leaving aside the two clear outliers (Michigan and California), these states have been in the center of educational debates, much like the top scorers: Alabama with high-profile spats among its state board of education,

Figure 2. Combined index of ESSA state plan ratings



legislature, and governor; Virginia stoutly resisting the Common Core; and Kansas in the thicket of school finance warfare. These states, with low-rated ESSA plans, typically have taken a combative approach with regard to federal education policy, and their low ratings may result from state policymakers assuming a defensive posture.

Second, there are notable outliers. Although the top-scoring group is less distinct, there are clear leaders and laggards. Washington and Tennessee easily received the highest honors, and Michigan and California are notable for their low scores. These states bear watching. Washington is notable as the only state for which ED had revoked an NCLB waiver, in 2014.¹⁶ The state had refused to adopt fixed teacher assessment in a way acceptable to the Obama administration, a deliberate approach that persisted in its fluid “menu of best practices” appearing in its ESSA plan.¹⁷ Yet, aside from teacher assessment, the state drafted a plan with high academic standards, clear expectations for schools and students, and clear guideposts for improving student achievement.



On the other hand, Michigan exemplifies a state plan with almost deliberate shirking. The state initially submitted what appeared to be an incomplete plan, a condition that Bellwether’s coders noticed with alarm. The state later revised the plan, only to have ED reject the revision as inconsistent with ESSA. In the process, the state’s governor and civil rights groups criticized the Michigan Department of Education, revealing the fractured arrangement of education governance in the Great Lakes State.¹⁸ Federal Secretary of Education Betsy DeVos eventually accepted a third draft but commented tersely that, while legal, the plan (from her home state!) was only a “baseline upon which to build.”¹⁹ We later provide more detail on Michigan’s case.

Correlates of ESSA Plan Content

Are there state characteristics that might begin to explain variations in ESSA plan ratings? We address that question by relating the ESSA ratings from Bellwether, A4ED, and Fordham to the four categories of factors we introduced earlier.

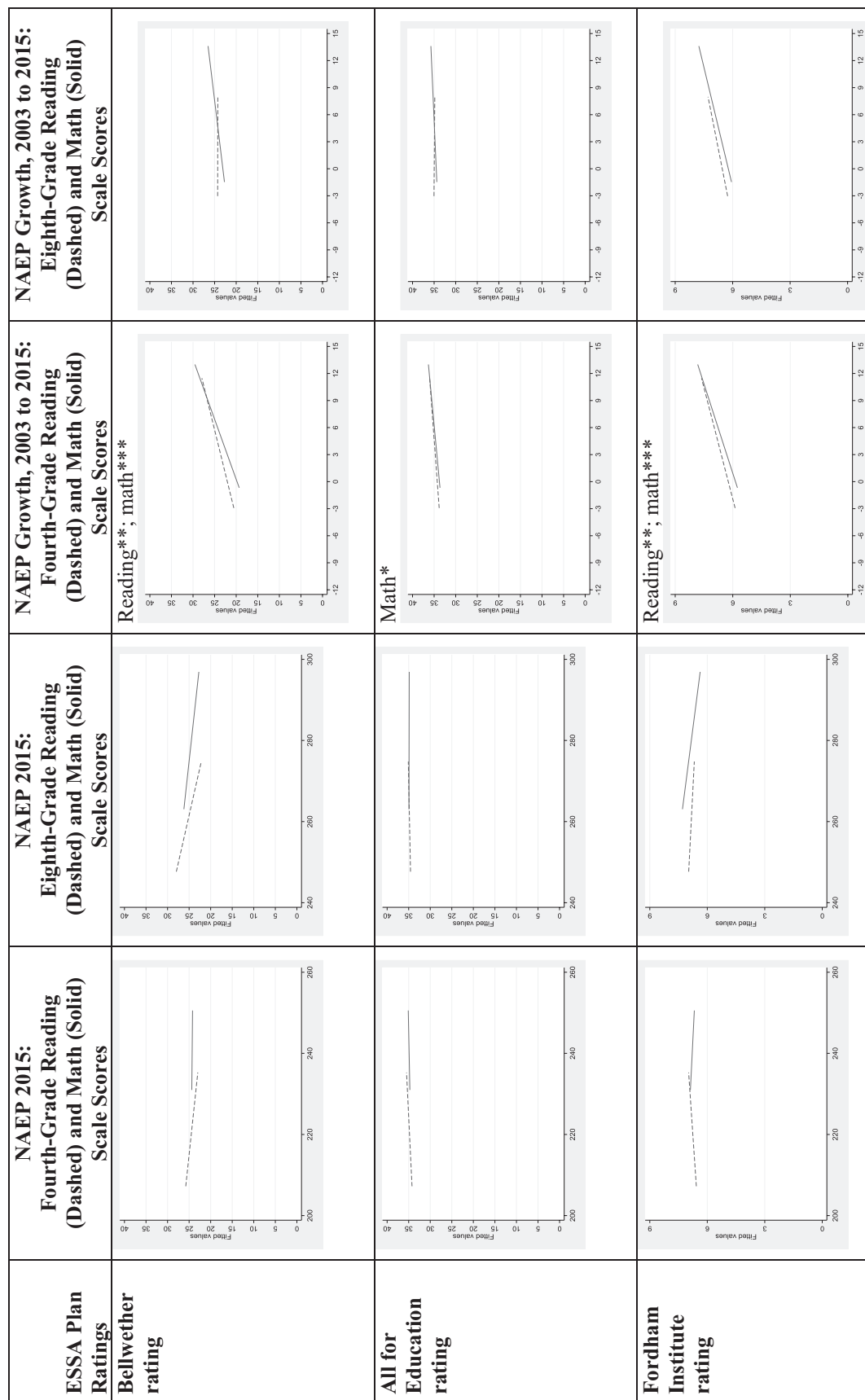
Prior student performance in states

“Prior student performance” refers to the pre-ESSA academic performance of a state’s students. Because states’ own measures vary, we use National Assessment of Educational Progress (NAEP) scores, focusing on scale scores to consider performance in fourth- and eighth-grade reading and math.²⁰ Additionally, we also use scale scores to measure gaps in achievement between more and less advantaged students in these same grades and subjects.²¹ The analysis includes a static and a dynamic approach to these scores. The static approach focuses on NAEP results published in 2015, the year that ESSA became law, whereas the dynamic approach considers the change in NAEP scores from 2003, the year after NCLB became law, to 2015.

Our first set of results appear in figure 3, which examines the static and dynamic results of fourth- and eighth-grade reading and math achievement. We have no strong priors here about what to expect, mindful that previously high-performing states might see strong accountability frameworks as key to academic excellence or might see other policy levers as important. Similarly, in poorly performing states, a weak achievement track record might provide motivation to tighten the accountability belt, responding to low performance by enhancing oversight, or it could suggest a loosening of the reins if policymakers see other methods (e.g., broader parental choice, reduced regulation, greater fiscal support) as more promising methods to foster improvement.

Figure 3. Relationship between ESSA plan ratings and state NAEP performance.

Note: Statistical significance of relationships is * $p < .10$; ** $p < .05$; *** $p < .01$



Overall, the results in figure 3 show an interesting contrast between our static and dynamic measures. The first two columns, which examine NAEP results from 2015, show no apparent relationships between prior achievement and the content of ESSA plans. Across the total of twelve bivariate relationships appearing in the plots (i.e., two lines in each of six plots), none was statistically significant at standard levels ($p < .10$ or better). In fact, the lines in each plot appear quite flat overall. Prior achievement measured in this static way, then, does not appear to help predict the level of accountability in state ESSA plans.

In contrast, the dynamic plots in the last two columns, in particular those in the third column, show detectable relationships. Increases in fourth-grade reading and fourth-grade math scores are both positively related to states receiving higher scores from Bellwether's raters and Fordham's raters. These results are statistically significant as well. Growth in NAEP achievement is also positively related to the A4ED ratings, but results are statistically significant only for fourth-grade math. The relationships in the fourth column, which show growth across eighth-grade reading and math, are either flat or positive, but none is statistically significant.

We consider another aspect of student performance in figure 4. Here we examine gaps in achievement between students not receiving meal subsidies and those receiving them. Interpreting this measure for the first two columns of the figure, the static measures, higher values represent larger achievement gaps between these groups in 2015. When we relate these measures to state ESSA plan ratings, we see in the first two columns that none of the twelve bivariate relationships are statistically significant at standard levels. The plot lines in the first two columns appear flat in this figure, suggesting no apparent relationships. The exception is the third row in figure 3, where there does appear to be a slight positive relationship with the Fordham ratings and achievement gaps. States with higher achievement gaps would be predicted to have higher scores from the Fordham ratings, suggesting a potential relationship between gaps and a desire to enhance accountability. We emphasize, however, that these results are only suggestive due to the low correlations and lack of statistical significance.

The same pattern of very low or no association with prior achievement holds in the last two columns of the figure, which examine changes in achievement gaps in fourth- and eighth-grade math and reading from 2003 to 2015. In these plots, upward-sloping lines would mean that growth in achievement gaps would be associated with higher ESSA plan scores and downward-sloping lines would show an association between states that had closed achievement gaps and had higher ESSA plan scores. Unlike the

Figure 4. Relationship between ESSA plan ratings and state NAEP scale score achievement gaps between students ineligible for meal subsidies and students eligible for meal subsidies.

Note: Statistical significance of relationships is * $p < .10$; ** $p < .05$; *** $p < .01$

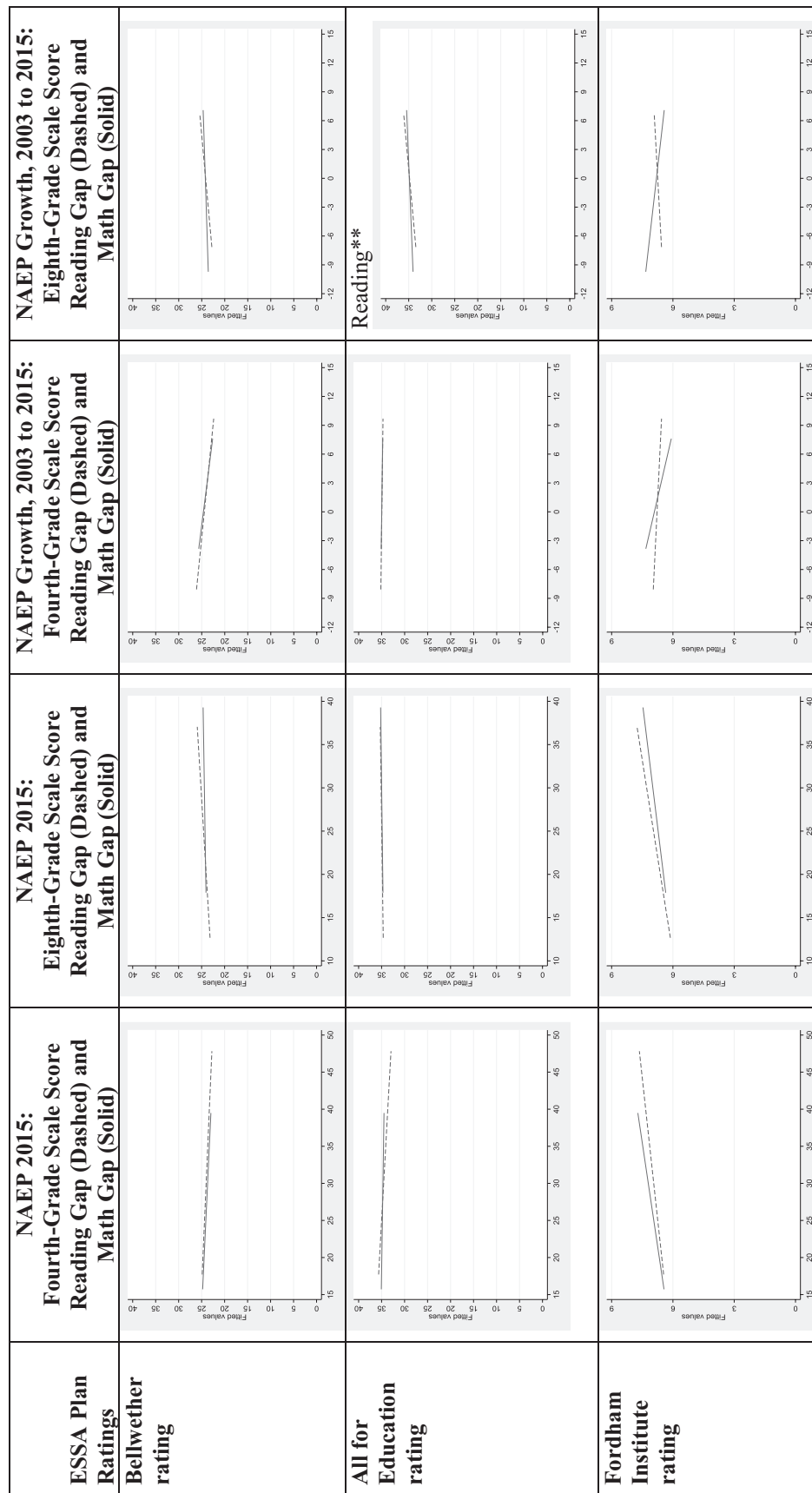


figure 3 results, showing some relationship between changes from 2003 to 2015 and achievement, there appears to be no consistent pattern to the relationships among achievement gaps in figure 4. Some lines are upward-sloping, some downward-sloping, and some essentially flat. Only one statistically significant relationship is apparent: growth in the eighth-grade achievement gap in reading is positively related to A4ED's scores.

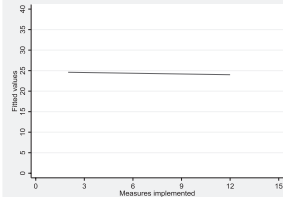
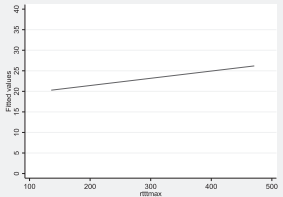
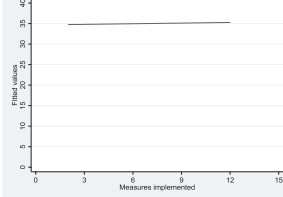
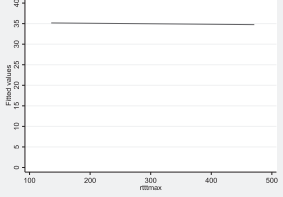
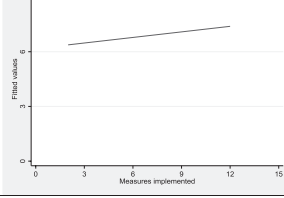
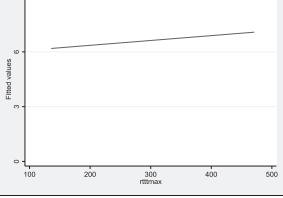
Prior state policy commitments

A second factor we consider as a possible factor explaining ESSA plan content is states' prior policy commitments in education. This factor allows us to examine whether states with previous commitments to accountability or educational innovation are likely to stay on that same path as they develop ESSA accountability provisions. We consider three factors here that march slowly back in time. The first is a measure from the organization Achieve, which surveyed states to determine which had adopted various measures, up to fifteen in all, to support college and career readiness in schools.²² These results are from 2013. The second measure, which examines policy from a few years prior to 2013, is the score that states earned in their Race to the Top (RTTT) grant application in 2010. On this measure, we chose for analysis the top score a state received when comparing its Phase 1 and Phase 2 RTTT proposals.²³ The third measure reaches back farther in time. Constructed by Thomas Dee and Brian Jacob, this measure indicates whether a state had a commitment to educational accountability in the years before 2002, which was the year NCLB became law.²⁴

Figure 5 presents results that, overall, reveal very weak or no associations between this prior policy content and the ESSA plan ratings from Bellwether, A4ED, and Fordham. The first column focuses on Achieve's college and career readiness measure. There is no relationship at all between it and the Bellwether or A4ED ratings, as the flat lines show. There is a suggestive positive relationship with Fordham's score, but the results there are not statistically significant. The second column presents the RTTT measure. Using that score to predict the ESSA plan ratings from Bellwether, A4ED, and Fordham shows a slight positive relationship between the RTTT scores and the ratings of Bellwether and Fordham, respectively, suggesting that high marks on RTTT might have signaled a future commitment to educational accountability in ESSA. Although suggestive, the relationships again fail to achieve statistical significance at conventional levels ($p < .10$ or better). Further, the A4ED results show no apparent relationship given the flat line in the plot. Our third measure takes the longer-term view by considering the presence or absence of accountability plans in states before 2002. To consider this factor, we

Figure 5. Relationship between ESSA plan ratings and previous state policy content.

Note: None of the relationships in the table is statistically significant at conventional levels.

ESSA Plan Ratings	Status of Implementation of Efforts to Support College and Career Readiness, 2013	Highest Score, Race to the Top Grant Contest, 2010	State Had Accountability System Pre-2002 (Yes or No)
Bellwether rating			Bellwether average Yes: 24.2 No: 24.4
All for Education rating			All for Education average Yes: 35.0 No: 34.9
Fordham Institute rating			Fordham average Yes: 6.7 No: 6.8

computed average ESSA plan ratings from Bellwether, A4ED, and Fordham by whether a state had adopted an accountability system in the pre-NCLB period. Those average scores appear in the third column of figure 5. They show essentially no differences at all, making a state's pre-2002 commitment to accountability a weak factor for predicting the content of state ESSA plans.

State implementation context

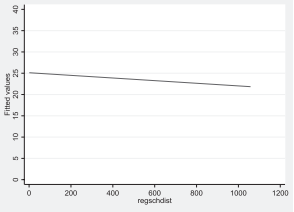
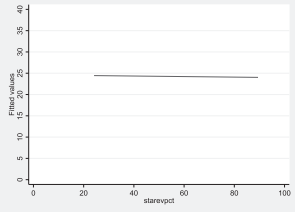
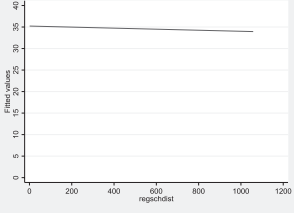
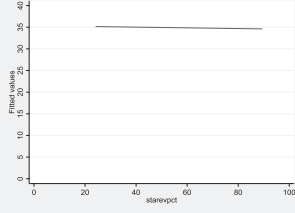
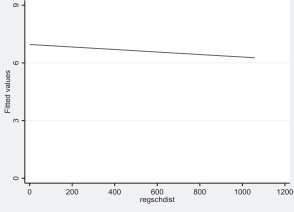
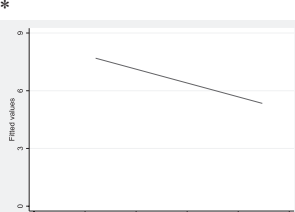
We also suspected that broader state contexts could influence how states envisioned implementing ESSA, and thus could inform overall state plans. We considered two aspects of state context. First was the overall K–12 governance context, which captured the broad institutional and financial arrangements that create a backdrop for education policy implementation in a state. The second contextual feature was the task demands facing educators across a state as they try to meet their students' educational needs.

Figure 6 presents the first part of these analyses of governance. The figure considers four different measures of the governance context: the number of school districts in a state, whether the state education chief is elected, whether state board of education



Figure 6. Relationship between ESSA plan ratings and state K-12 education governance context.

Note: Statistical significance of relationship is * $p < .10$; ** $p < .05$; *** $p < .01$

ESSA Plan Ratings	Number of School Districts in State	State Education Chief Is Elected	State Board of Education Members Are Elected	Percentage of K-12 Revenues from State
Bellwether rating		Bellwether average Elected: 24.6 Not elected: 24.2	Bellwether average*** Elected: 17.7 Not elected: 25.3	
All for Education rating		All for Education average** Elected: 33.1 Not elected: 35.6	All for Education average Elected: 35.3 Not elected: 34.9	
Fordham Institute rating		Fordham average Elected: 6.4 Not elected: 6.9	Fordham average Elected: 6.7 Not elected: 6.8	

members are elected, and the percentage of revenues for K-12 education from the states. We suspected that states with more centralized systems of governance (i.e., fewer school districts; higher responsibility for K-12 funding) would be more likely to demand greater accountability. Further, we expected states with elected state education agency heads and state board members to be less likely to demand more accountability given the poor experiences that many states had with NCLB implementation and the potential pressure that organized interests and voters can put on elected officials to give local communities more discretion for implementing ESSA.

The figure 6 results partly support our expectations. The first column is only suggestive (given that results were not statistically significant) that states with more school districts received lower marks from Bellwether, A4ED, and Fordham. On selection of state officials—the middle two columns of figure 6—two of the six relationships we examined were statistically significant and substantively meaningful in the way that we had expected. On the Bellwether rating, states with elected board members scored on average 7.6 points lower than states with nonelected boards ($p < .01$). Similarly, on the

A4ED scale, states with elected chiefs scored 2.5 points lower than states with nonelected chiefs ($p < .05$). The remaining four comparisons did not reveal statistically significant differences, and in most cases the differences were less than one point in the various comparisons. Finally, the third column of figure 6 shows that financial governance is negatively correlated with ESSA scores from the three organizations—counter to our expectations, given that we had expected states footing more of the K–12 bill to demand greater accountability. The relationship is weak for the Bellwether and A4ED ratings, but it is statistically significant ($p = .054$) for the Fordham rating.

Our final look at state contexts focuses on the task demands that local school districts and educators likely face as they teach their students. We use two indirect measures to infer these task demands. The first is an index of challenging student populations that we constructed by assembling information on the proportion of students in each state who are students of color, eligible for subsidized meals, learning English as a second language, and receiving services under the Individuals with Disabilities Education Act.²⁵ The second indirect measure, an overall average of state poverty rates from 2014 to 2016, allowed us to capture the general level of resources one might expect in students' homes that would augment school efforts to support student learning. Based on evidence from previous research, we expected states with more challenging populations and states with higher poverty rates would have lower accountability expectations.²⁶

The results in figure 7 challenge our expectations. The strongest relationships existed for the Bellwether ratings (first row of the figure), where we see that states with students possessing greater needs and states with higher levels of poverty were more likely to earn high scores from the Bellwether raters, and both results were statistically significant at $p < .10$. There also was a positive relationship between these two measures of task demands and the ratings from A4ED and Fordham, although the relationships were weaker and not statistically significant. We further probed the Bellwether results by breaking apart our student index into its constituent parts and found that the main element of the index driving the results was the percentage of students in each state eligible for subsidized meals, which was related to the Bellwether scoring at $p < .10$. None of the other constituent parts of the index was statistically significant.

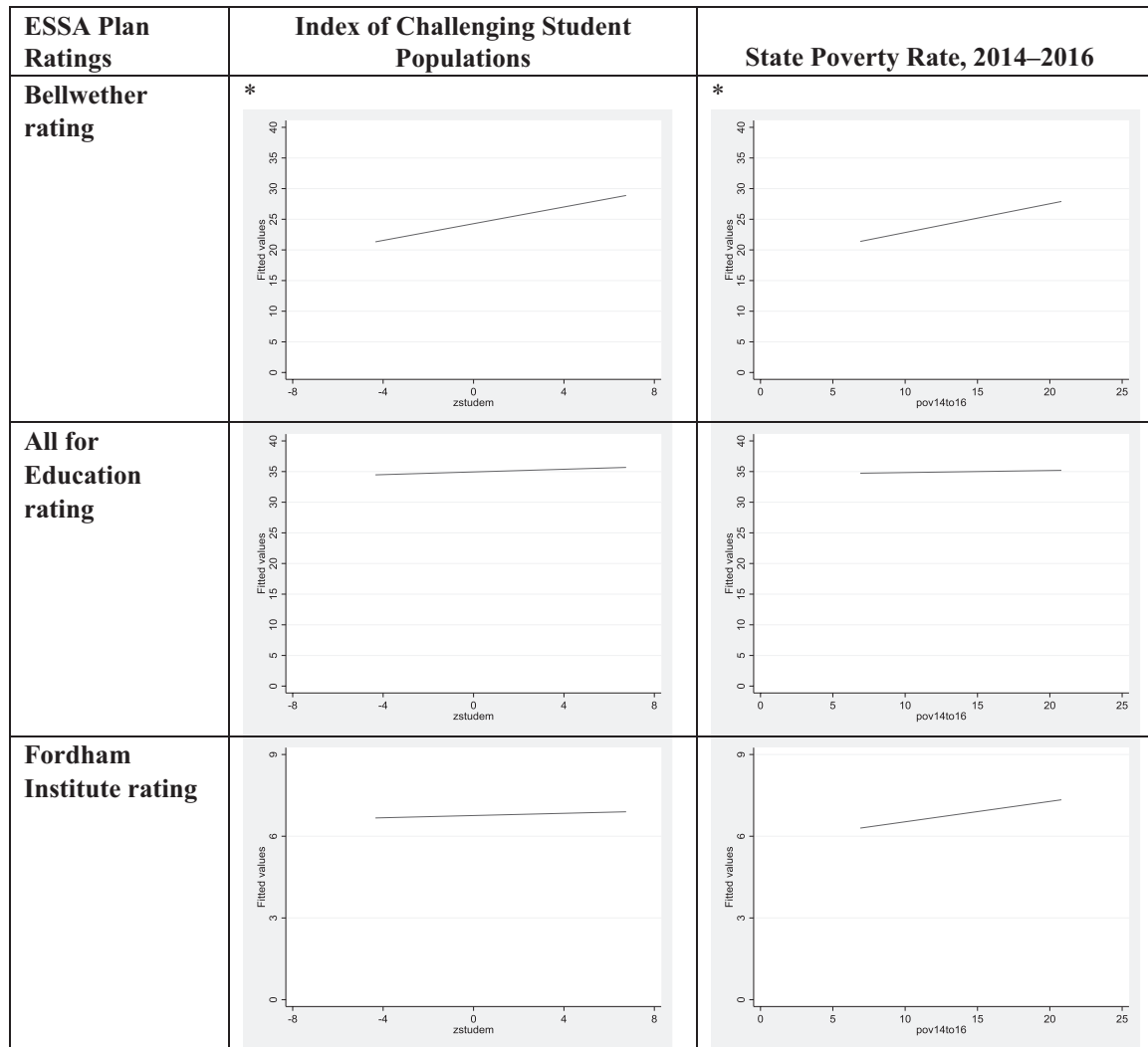
State political context

The final set of factors we considered to predict state ESSA plan scores focused on politics. The first was a measure of party control of state government, which we examined via two different comparisons.²⁷ In the first, we compared plan scores by



Figure 7. Relationship between ESSA plan ratings and educational task demands.

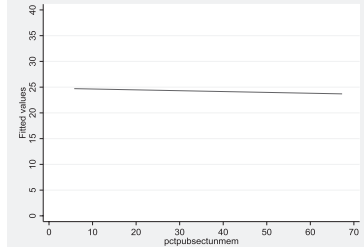
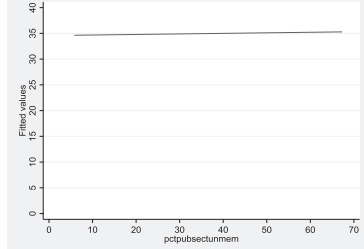
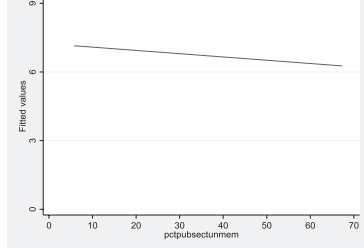
Note: Statistical significance of relationships is * $p < .10$; ** $p < .05$; *** $p < .01$



reviewing whether, after the 2016 elections, the same political party controlled both a state's governor's office and its legislature, known as unified government, or whether split party control, known as divided government, prevailed. A second comparison focused on those states where control was unified: we compared state ESSA plan scores in the seven states with unified Democratic control with the twenty-four that had unified Republican control. The second overall measure of state politics was a measure of the percentage of public sector workers in each state in 2016 who were members of labor unions.²⁸ We use this measure because teacher unions have traditionally been a strong force in state education politics. Unfortunately, teacher union membership is

Figure 8. Relationship between ESSA plan ratings and state political context.

Note: Statistical significance of relationships is * $p < .10$; ** $p < .05$; *** $p < .01$

ESSA Plan Ratings	Party Control: Divided vs. Unified	Republican vs. Democratic Control	Percentage of Public-Sector Workforce Who Are Union Members
Bellwether rating	Bellwether average Divided: 25.0 Unified: 24.1	Bellwether average Republican: 23.7 Democrat: 25.4	
All for Education rating	All for Education average** Divided: 35.9 Unified: 34.2	All for Education average Republican: 34.1 Democrat: 34.3	
Fordham Institute rating	Fordham average Divided: 6.4 Unified: 6.9	Fordham average Republican: 6.0 Democrat: 7.2	

not publicly reported across all unions and states, yet given that teachers compose a sizable portion of public sector workers in each state, a measure of overall public sector unionization rates is a reasonable substitute.

The relationships between our measures of state politics and ESSA plan ratings appear in figure 8. Overall, the relationships are mainly suggestive or not apparent. On division of partisan control, appearing in the first column, the results show inconsistent relationships. The A4ED results are the only ones that show a statistically significant difference. On average, states with divided party control had higher scores, by 1.7 points, on the A4ED measure compared to states with unified party control. The second column shows that states with unified Democratic government had higher scores across the Bellwether, A4ED, and Fordham ratings compared to states with unified Republican government. The results are merely suggestive, though, because none of



those differences was statistically significant. Finally, the third column presents plots showing the unionization measure. The results show no apparent relationship between the percentage of public sector employees who are union members and state ESSA plan scores. The Bellwether and A4ED ratings are essentially flat, whereas the Fordham plot shows a slightly negative, albeit statistically insignificant, relationship.

State Case Studies

Although each state entered the ESSA world with its own history with NCLB and RTTT, we selected five cases to illustrate the range of political and policy responses states took: Michigan, New Mexico, Iowa, California, and Florida. Although these states do not cover every political tussle that surrounded ESSA's rollout, they are representative of the approaches that states across the country took. In the cases here, we consulted Bellwether's evaluations of state ESSA plans to help identify key parts of state plans, as Bellwether's reviewers were the most thorough and specific of the three evaluators we considered. However, we also directly used the state plans (original and revised), local media, and ED's communications in drawing these vignettes.

These cases have the further advantage that they cover a range of political governance types—a factor we expect plays a significant role in explaining differences in ESSA approaches.²⁹ The five states at issue here illustrate almost the full range of selection mechanisms for state schools chiefs and state boards, as shown in table 1. State boards are selected by either the governor or voters, and state schools chiefs are selected by either the governor, voters, or the state board. These arrangements may help to explain both the consistency of state ESSA plans as well as interinstitutional conflict. Due to the differences in electorates, states with strong gubernatorial presence can expect the support of state resources in a way that voter-dominated systems cannot.

Of the independent evaluators groups we surveyed, Michigan had the most consistently low-rated plan. That plan—and its political aftermath—suggests a state with little coherent academic vision. At the other end of the scale, New Mexico's plan took the process seriously and comprehensively, both in its accountability focus and its extensive stakeholder outreach. This case also shows the perils of policy that is closely identified with state political figures, as a change in state leadership undid half of New Mexico's initial efforts. Iowa represents a midway case. The state long has had difficulty embracing accountability-driven education in the face of immense local control, despite a deep tradition of measuring educational progress. The raters here placed Iowa near the middle of the pack.

Table 1. State Education Governance Arrangements

<i>State</i>	<i>State Schools Chief's Path to Power</i>	<i>State Board of Education Members' Path to Power</i>
Iowa	Governor appoints	Governor appoints
Florida	State board appoints	Governor appoints
California	Voters elect	Governor appoints
New Mexico	Governor appoints	Voters elect*
Michigan	State board appoints	Voters elect

* New Mexico's board is an advisory body.

Finally, California and Florida represent extremes of a different nature. Not only are both states large (and jointly home to 15 percent of all K–12 students), but they have longstanding state accountability systems that do not conform to the federal “ideal.” With ESSA, both states fought to retain distinctive elements of their state systems, and California and Florida were the last two states with approved ESSA plans. In addition, California and Florida have taken opposite roads on accountability: Florida’s system emphasizes a single A–F metric with specified consequences, while California long has had “dashboard”-style accountability that does not rank schools.

Iowa’s ESSA plan

ESSA offered Iowa a chance for a fresh start at federal accountability policies. Iowa was one of the few states to have an NCLB waiver denied outright, in part because the state maintained a strong ethos of local control that both NCLB and RTTT challenged. Indeed, ED noted that the waiver was denied in 2012 not only because of the absence of student achievement in its evaluation, but also because the Iowa Department of Education did not “have the authority to ensure implementation” in any event.³⁰ Yet, by 2016, Iowa sought to incorporate both local concerns through an extensive outreach program and align with ESSA’s somewhat softer academic strictures. Unlike in 2012, Iowa’s ESSA plan drew praise from multiple observers, and ED did approve a revised version of the state’s plan. (Iowa was the forty-fourth plan that ED approved.)

Iowa’s ESSA plan drew praise from Bellwether and, notably, the Iowa Department of Education even featured its review on its ESSA plan development page.³¹ Bellwether was particularly impressed by the state’s combination of outreach efforts, “compelling” school indicators, and “positive signal[ing] to schools.”³²



Iowa's ESSA plan was blunt about its stakeholder and expert input, as its *primary* guiding principle was "implementing an inclusive process . . . that demonstrates value for our partnerships with these stakeholders."³³ Like New Mexico, Iowa took the requirement to engage stakeholders quite seriously and produced well over one thousand stakeholder comments from thirty-one listening sessions, forums, and public meetings held in eighteen locations around the state, as well as from online sources. Iowa also included a seventy-page description of its efforts to include stakeholders and the general public—far more than most other states.

Bellwether praised the state's efforts as a "smart process" and highlighted Iowa's plan to continue stakeholder and expert input as ESSA came online. This was especially true because Iowa's accountability system was in transition and expert input could prove invaluable in bringing that system up to speed. This feeling was seemingly widespread, as many participants in the listening sessions sought assurances from Iowa officials that schools would be allowed to give input on future work and would retain as much local flexibility as possible.

Whatever the stakeholder input, Bellwether was enchanted by Iowa's "honest and direct analyses" of the challenges of closing achievement gaps and its use of similar metrics for both elementary and high schools. Bellwether argued that this created aligned accountability across levels. Throughout its report, Bellwether noted that Iowa was a model for explaining *why* the state took particular approaches. The state's indicators included student participation rates, students' growth, and progress toward English-language proficiency, as well as a student survey of engagement, safety, and learning environment. Bellwether noted that Iowa was among the few states to focus adequately on schools with the most academic trouble. Others praised Iowa's emphasis on its multitiered system of supports to reduce achievements gaps.³⁴

Still, in December 2017, ED returned Iowa's initial plan and asked for revisions—and, ironically, took issue with some of the same elements for which Bellwether had praised the state. In particular, ED argued that several of Iowa's proposed academic indicators did not meet ESSA's requirements. (Iowa was not unique in this regard—ED prompted many states to revise their indicators.) These included the graduation rate indicator, the English-language proficiency indicator, and the student engagement survey. In each case, ED wrote that Iowa's plan either had too little detail or inappropriately combined elementary and high school data. It further faulted the state for gathering too little stakeholder input on future assessment issues, as well as being too optimistic

about English-language learner (ELL) proficiency. (ED noted that Iowa needed to use *progress* toward proficiency rather than actual proficiency.) But ED's biggest issue was the state's attempt to remove low-performing schools from supports when they showed improvement but remained in the lowest 5 percent statewide. Nevertheless, these ED challenges did little to Iowa's basic stakeholder-centric framework, and ED approved the plan in May 2018.

New Mexico's ESSA plan

New Mexico has long struggled with lagging academic achievement, despite the high-profile efforts (and controversy) of former Governors Bill Richardson and Susana Martínez. The state made efforts to have its plan be among the first ESSA plans approved in order to create “stability and continuity” for New Mexico schools, according to the acting secretary of the Public Education Department (PED).³⁵ Yet New Mexico's case illustrates the risks to federal incentives from high-profile state political battles. After the 2018 elections, the state substantially changed its ESSA plan.

The state's initial plan received among the highest marks from Bellwether and other groups, and it was developed with an explicit focus on stakeholder input. PED visited six cities twice to collect feedback on the department's ESSA plans. Bellwether noted that, in addition to measurement and identification, New Mexico had an “aggressive, concrete list of interventions” and was “taking seriously the challenges faced by those schools.”³⁶ Bellwether's primary concern about New Mexico was nebulous language describing when schools would no longer need additional state support—that is, it was unclear when schools' students would have sufficiently improved. However, ED was sanguine. It was particularly impressed by the plan's clear measurements for each ESSA subgroup, the A–F school-based grading system used to identify schools for supports, and the provision of teacher and principal mentors for these identified schools.

This did not prevent a sea change in New Mexico's education politics. In the 2018 gubernatorial campaign, Michelle Lujan Grisham ran against major features in the ESSA plan, tying those features to the outgoing (state) Republican administration. Instead, she promised to “shift the philosophy . . . [to] celebrating success” by revoking the A–F accountability system, the PARCC tests, and teacher evaluation.³⁷ Her campaign tapped into a reservoir of distrust among a portion of the state's education establishment, including the chairwoman of the New Mexico legislature's Legislative Education Study Committee. She told the *Albuquerque Journal* in no uncertain terms that the ESSA plan was poor. She said that outside groups such as Bellwether were ill



suited to comment on New Mexico's needs. The plan itself engendered "hatred and dislike," she said, especially for its pursuit of teacher evaluations and A–F grading.³⁸

Lujan Grisham won the election, and shortly after taking office she and PED directed low-performing schools to stop implementing ESSA turnaround plans.³⁹ PED immediately sought to replace New Mexico's school accountability and interventions by jettisoning PARCC's name (although the tests themselves remained similar), elements of direct student services, and ESSA's "more rigorous interventions" (MRI) in low-performing schools. In their place, PED required curricular reforms and "community schools models." For assessment, the department emphasized educational climate and student growth rather than academic proficiency. The department, with the legislature's aid, eliminated the single summative A–F system with a "family-friendly" dashboard of information. Some elements remained, however, including the indicators for measuring performance and school standards. Nevertheless, despite the state's retreat from academic accountability (and the legal ambiguity of its initial revisions), the federal government accepted the changes in July 2019.⁴⁰

Michigan's ESSA plan

Michigan's plan was widely panned by independent evaluators as among the weakest responses to ESSA. EdTrust noted bluntly that "in little more than a decade, Michigan has gone from being a fairly average state to among the nation's bottom ten."⁴¹ The organization faulted the state's plan for being dishonest, unambitious, and misdirected. Bellwether likewise lamented that Michigan's plan seemed incomplete. Both the state's governor and lieutenant governor urged Secretary DeVos to demand changes from the Michigan Department of Education (MDE), especially with regard to special education students.⁴² This disconnect is clearly a result of divided governance—neither the governor nor the legislature have a say in the state's education leadership, even though they control the education department's budget.

The governor and lieutenant governor may have been right. Shortly after Congress enacted ESSA, MDE noted that it would try to repurpose the state's existing reform effort, "Top 10 in 10," in service to ESSA. MDE officials wrote that doing anything else would "create a duplicative structure."⁴³ Although Michigan's efforts were among the most robust in the country—more than 350 parents, stakeholders, scholars, and politicians participated—their comments were not directly specifically at ESSA implementation, signaling Michigan's lack of investment in carrying out the law.

In line with Bellwether’s observations, ED returned Michigan’s ESSA plan with a demand for more detail. The department called the plan “insufficient . . . to review” and said that ED needed substantially more “clarifying information.” The department said the state had not indicated how it would measure student learning, support migratory children’s learning, or spend federal teacher support funds, or said whether it would continue to consult with teachers and other required stakeholders during implementation. It also took the lieutenant governor’s demands to heart, and ED required Michigan to address equitable access to education for special needs children before approval.⁴⁴

MDE was unimpressed, and the state superintendent suggested that ED wanted only “small, nitpicky stuff.”⁴⁵ Michigan did provide some details for ED, but with minimal explanation or clarification. Instead of robust accountability number measures, the state opted for a dashboard. Instead of concrete goals, the state set back-of-the-envelope projections. Even the MDE’s press release was unapologetic: “We appreciate those individuals and groups who were genuine and supportive . . . there were outliers who had different opinions, and they were heard,” State Superintendent Brian Whiston said.⁴⁶

Independent reviewers at Bellwether questioned the feasibility of the state’s overall goal of becoming a “top 10” state, given that Michigan did not indicate how it would measure progress toward that goal or whether that goal was feasible given existing student performance. Reviews also faulted the state for gaping holes in its implementation strategy, especially around academic measures. There was “no substantive detail about how these indicators are valid, reliable, backed by research, increase student learning, relate to post-secondary outcomes, or meaningfully differentiate schools,” Bellwether’s reviewers said.⁴⁷

Michigan did address ED’s concerns, but only just. In November 2017, DeVos approved the plan with a terse statement that Michigan’s plan was in compliance with the letter of the law and not the spirit. “While the plan meets the statutory requirements, Michigan must not view this as a ceiling, but rather as a baseline upon which to build,” she said.⁴⁸

As in New Mexico’s case, Michigan’s ESSA work ran into state political controversy. Not only had its ESSA work been faulted by the state’s chief executives, but the state legislature faulted MDE and the state board for their ESSA response. Key critics



echoed ED to argue that the board's approach was "nebulous" and "mask[ed] poor performance."⁴⁹ When MDE did not change course by December 2018, the legislature enacted an A–F accountability system *in addition* to the ESSA dashboard.

Due to differences in coverage, Michigan's A–F plan may not be in compliance with ESSA requirements, according to the state's interim state superintendent, but the law did not require replacing the existing metric. Instead, MDE was supposed to report a second, parallel measure—something along the lines of Florida. As of early March 2020, however, MDE still had not issued any school grades.

California's ESSA plan

California never took kindly to NCLB, and insisted on maintaining a state accountability system in parallel with a federal accountability system. ESSA offered the state an opportunity to build a single system that was acceptable to both state leaders and the federal government. Although the state still resisted complying with ESSA, it was able to accommodate ED's requirements.

California's initial ESSA plan contained a strong commitment to robust curriculum standard and assessment instrument—but a strikingly limited vision for how to encourage students and teachers to use it. The state used a set of standards based on the Smarter-Balanced initiative, a new test for science standards, and a well-regarded language proficiency exam. It also incorporated generally used indicators of student and school progress, including changes in test scores, chronic absenteeism, suspension rates, and completing college or career work (IB exams, AP tests, college courses, and so on). Yet in its initial response to California, ED noted that the state's plan failed to include long-term goals for students' performance, graduation rates, or English-language proficiency, or a way to incorporate students' test scores into an assessment framework.⁵⁰

California resisted using these metrics directly as assessment tools, and proposed that it be allowed to use a set of holistic "dashboard" indicators. State Board President Michael Kirst argued that any single number would be misleading: "It's sort of like using the Dow Jones average for everything in the stock market. No reasonable investor would do that."⁵¹ Kirst also suggested that its existing twenty-five-box matrix provided better information to parents, as they could identify the metrics most important to them. Nevertheless, outside raters, columnists at the *Los Angeles Times* and *Sacramento Bee*, and ED all found the system convoluted. Betsy DeVos said that the state "took a simple concept like a color-coded dashboard and managed to make it nearly indecipherable."⁵²

Despite DeVos's criticism, ED did not object to the dashboard as much as its lack of coordinated metrics. Although Kirst had argued that weighting individual measures to create a single score was arbitrary, ED pushed the state to create specific, standard-based measures weighted for their relative importance (as required by ESSA). The department particularly objected to the state's relative measuring scheme. ED inquired whether it would be "possible that a school might exit [receiving additional targeted support] because other schools began to perform worse."⁵³ Further, the state only proposed to use overall school measures. Such an approach would allow California to measure overall cohort changes but would prevent the state from knowing whether any particular groups of students were making genuine gains. Even after the state filed a revised plan, ED objected to California's use of "approximate status" as an indicator of student progress.⁵⁴ It was clear that California wanted to resist identifying schools as much as possible, but the federal government wanted specificity.

For the federal government, however, the biggest disconnect was between its insistence on identifying schools for support and California's desire to support districts—in part stemming from the state's top-down school finance system. ED wanted the state to identify the bottom 5 percent of school performers—a single measure ill suited to a dashboard approach. In the end, California promised to identify schools that were either "poorly" or "very poorly" performing on half or more of the dashboard indicators. This strategy could identify somewhat more than 5 percent of schools, but it did not necessitate revamping the state's accountability system or its commitment to the dashboard. ED accepted the proposed changes and approved the plan—the penultimate state (tied with Utah) to receive DeVos's approval.

The California Department of Education explicitly highlighted the differences between California's model and the federal vision. California's plan primarily monitored school districts rather than schools, but California noted that "a compromise was reached [with ED] that allows the state to continue working with districts, recognizing that the challenges of individual schools are often related to wider systemic problems." State Board President Michael W. Kirst was likewise optimistic: "Given the differences between state and federal law, the plan approved by Secretary DeVos . . . represents the best possible outcome of our discussions with U.S. Department of Education staff," he said.⁵⁵ California's position was strengthened because state work was supported both by the governor (through the state board) and by voters (through the state chief). ED could expect that the California Department of Education would represent something of a state consensus on education policy.



Florida's ESSA plan

Florida was the final state whose ESSA plan was approved by ED. As with California, the state had strong preferences about its approach to educational accountability, and it was similarly reluctant to bend to the strictures of ESSA. To some extent, Florida never did, as it maintained a separate accountability metric for federal purposes that was branded the “Federal Index.” In distinct contrast with many other states, Florida did *not* engage in widespread outreach to write the plan, but the state argued that relying on state open-meetings laws sufficed for public input. In its proposal, Florida noted that these routine meetings were “attended by education stakeholders (including teachers, principals, other school leaders, parents, and other stakeholders).”⁵⁶

Florida’s A–F accountability system dated from 1999. The state legislature has a strong track record of dictating various components of the K–12 system, meaning that the state’s education bureaucracy was more constrained in its responses to the federal government than those of some other states. When NCLB became law, the state essentially operated with two distinct accountability systems. In communications about school performance, state officials preferred promoting discussions along the lines of their more graduated system rather than the more blunt “met or did not meet AYP” that NCLB required.

Although some outside reviewers saw Florida’s initial proposal as a promising plan due to its clear accountability and curriculum focus, others noted the state’s obvious avoidance of subgroup scoring and its deliberate choice to test only in English. It was “alarming” that Florida’s plan provided no accommodations for ELLs, and the plan also could deprive schools with low-performing student groups of targeted support, according to Bellwether Education Partners’ review.⁵⁷ ED shared these concerns, which singled out the state’s lack of ELLs in its accountability metric, failure to split out scores for subgroups of students, and limited to no accommodations for providing tests in languages other than English. Still, those comments did emerge in a context where Secretary DeVos remained generally enthusiastic about and willing to share praise for Florida’s past history of academic gains for students.

The Florida Department of Education (FLDOE) defended its approach, arguing that *all* students were better served if schools were held accountable for *all* students’ performance. In its A–F system, the state identified schools based on the performance of the lowest-performing 25 percent of students, regardless of those students’ other subgroup memberships. The federal government objected, with full support from some

state civil rights groups. The state resisted, and FLDOE toyed with requesting a waiver from this portion of ESSA in July 2017.

That action angered the civil rights groups, especially given the state's history with ELLs. Since August 1990, FLDOE had operated under a consent decree to ensure that ELL students had equal access to educational programs. The groups argued that the state's big-basket approach would all but guarantee that ELL students would be left out. "It's unconscionable that there would be no accountability applied at the school level for this very primary function . . . especially in an immigrant-rich state such as Florida," a South Florida civil rights activist said.⁵⁸

The waiver was never filed, but the state submitted a proposal in September 2017 that appeared to contain large elements of the proposed waiver's text. ED promptly rejected the proposal, and in early 2018 Florida acquiesced to a separate federal index that included ELL proficiency and agreed to separate out individual subgroups. Still, the state also maintained that its academic performance was well above average based on national rankings, whatever the system. The state was bolstered by none other than Secretary DeVos, who, as noted earlier, publicly praised Florida as a "bright spot" for "significant improvement" in NAEP test scores. It seemed clear that Florida had little to fear once the technical requirements of ESSA were met.⁵⁹

By late summer, Florida was preparing to submit a final set of ESSA revisions. Governor Rick Scott maintained that Florida did not need federal prodding: "Our record of educational excellence and strong performance should serve as a testament that our system is working and that our state understands how to best serve our students," he said.⁶⁰ The final set of changes seemed sufficient, and, after five attempts, ED signed off on Florida's plan in September 2018.

Discussion and Conclusions

What can we make of states' responses to ESSA? In many ways, ESSA was a reaction to exhaustion with the approaches and assumptions embedded in NCLB, the RTTT grant competition, and NCLB waivers. Although ESSA maintains the possibility for higher standards and better monitoring of poor academic performance, judging whether states have taken up that call remains difficult to assess. The Bellwether, A4ED, and Fordham ratings that we examined in detail provide excellent ways to probe state initiatives to see how they are operating in the new ESSA era.



Whether these state approaches have embodied the letter and spirit of ESSA is a matter of judgment, given the many different programs that the law encompasses and the varying conceptions of its purpose. From one perspective, states adopting relatively diverse approaches (some even backing away from what might be considered a higher bar for accountability) is no problem at all and might even be judged a success, given that one interpretation of ESSA's mandate was simply to unleash the states to try different strategies. That perspective sees ESSA, and in particular its Title I provisions, mainly as a funding stream designed to help top off state and local education budgets. A different perspective sees ESSA as a more robust federal education program with particular goals that embrace a particular vision of how states should go about accomplishing them. Those endorsing the latter point of view would be disappointed at the significant variation across state approaches to accountability, and in some cases state resistance to federal preferences.

Looking at our quantitative results and our case studies, it is clear that the states have sought to squeeze as much variation out of ESSA as possible. It appears that political emphasis on academic performance has receded, yet our results showing that growth in achievement in fourth grade (but not eighth grade) did help predict the content of state ESSA provides one important data point in the other direction. Overall, it may be that accountability is now more of a millstone to politicians than a motivator of schools. These patterns suggest that many in the states are glad to embrace the first perspective on ESSA that we describe above, seeing it as a way to enhance revenues so states can pursue their own paths and theories of action that are intended to improve educational outcomes. This is especially clear in the cases of Michigan, New Mexico, and California, and in other states as well. This line of thinking also provides one interpretation of our finding that states with needier school populations received higher scores from the ESSA plan raters. Civil rights and other groups (such as the National Urban League) advocated for ESSA precisely because it could funnel more resources into needy schools. If that impulse is dominating in the states—and more will become clear as state ESSA plans move deeper into implementation—it could be that the tide of performance is out at the moment. A more optimistic take is that states with needier student populations and more poverty believe that developing accountability measures in ESSA remains an important priority and a way to enhance educational equity and excellence for all students.

If the pessimistic view is correct, it might help explain why persistently low-performing states lack the political will and desire to emulate better performers. After

thirty years of beating the accountability drum, there simply may be little political gain for contemporary politicians to continue playing off the same sheet of music. Instead, state governments have to fight rear-guard actions justifying the length of accountability measures and preventing parents from nixing their children's test participation. To date, states have won these battles, but these are not fights that win votes. Our finding of a lack of consistent relationships between the ESSA plan ratings and different features of state politics also suggests that, unlike other issues, the politics of education is not nationalized to the extent that other issues, such as health care and gun control, are. Episodic and contextual politics seem to dominate at the state level, as our quantitative results and our case studies appear to show.

In terms of policy development, the fact that our results showed the content of prior policy in the states to be a poor predictor of ESSA plan content suggests a potentially concerning path forward, at least when one considers the states as a whole. When past policy only weakly predicts contemporary policy developments, it suggests that policy development in education is more of a random walk than a process of experimentation, learning, and systematic revision. Given the increasing challenges facing the nation's schools due to the expanding diversity of their students, widening overall inequality in society, and the contemporary crisis that the COVID-19 pandemic is posing in state and local education systems, it likely would be reassuring to know that the states are operating in a more reflective way.

Yet perhaps there remains a silver lining. The diversity of state plans fulfills the central premise of ESSA: let the states take the lead. If states fail to help their students reach their potential—whether that is measured by attendance, academics, or other indicators—then state leaders can no longer blame federal law. The national tide of reform is out, but when it returns, high-performing ESSA states will be well positioned to show that states, or at least *some* states, do have the political capital, educational expertise, and institutional capacity to succeed. In the run-up to NCLB, both North Carolina and Texas served this role, and that law borrowed heavily from those states' work. If ESSA's flexibility leaves low-income and less-advantaged children worse off in five, ten, or fifteen years, high-performing states that are better at meeting their needs in particular and the needs of students across the economic and achievement spectrum in general likely will again chart the future, given that federal reform directions typically are a lagging indicator of state efforts already underway. Whether that bodes well for all students will depend on state policymakers and advocates in the interest group community systematically learning about what it will take to improve



policy and, subsequently, whether leaders will be able to muster the political will to make those improvements. In that regard, although the ESSA era has reshaped the playing field for federal, state, and local officials, the key policy learning and political challenges that have confronted the states since the first ESEA became law more than fifty years ago remain highly salient today.

ACKNOWLEDGMENTS

The authors would like to thank Chester E. Finn, Jr., Christopher Ruskowski, Andy Smarick, and other HESI members for helpful comments on earlier drafts, and also personnel in state departments of education who read the case studies and provided feedback. Any errors of interpretation that remain are our own.

NOTES

- 1 White House Office of the Press Secretary, “Remarks by the President at Every Student Succeeds Signing Ceremony,” December 10, 2015, <https://obamawhitehouse.archives.gov/the-press-office/2015/12/10/remarks-president-every-student-succeeds-act-signing-ceremony>.
- 2 Rebecca R. Skinner, *The Every Student Succeeds Act (ESSA) and ESEA Reauthorization: Summary of Selected Key Issues*, Congressional Research Service, Washington, DC, December 7, 2015.
- 3 A great debate over the pros and cons of the ESSA approach, as well as historical perspectives on ESEA’s development since 1965, appears among the authors in Frederick M. Hess and Max Eden, eds., *The Every Student Succeeds Act: What It Means for Schools, Systems, and States* (Cambridge, MA: Harvard Education Press, 2017). In the same volume, see Chester E. Finn Jr., “Fifty Fast Years.”
- 4 See Paul Manna, *School’s In: Federalism and the National Education Agenda* (Washington, DC: Georgetown University Press, 2006), on state goals in the 1980s.
- 5 Marguerite Roza, *Educational Economics: Where Do School Funds Go?* (Washington, DC: Urban Institute, 2010).
- 6 Daarel Burnette, “State Chief Pledge to Continue Focus on Accountability,” *Education Week*, November 20, 2015, https://blogs.edweek.org/edweek/state_edwatch/2015/11/state_chiefs_pledge_to_continue_focus_on_accountability.html.
- 7 National Governors Association, “ESSA Implementation [Letter to Secretary of Education John B. King Jr.],” April 4, 2016, www.nga.org/letters-nga/essa-implementation.
- 8 National Association of State Boards of Education, “NASBE Issues Statement on Final ESSA Accountability Regulations,” press release, November 28, 2016, www.nasbe.org/press-releases/nasbe-issues-statement-on-final-essa-accountability-regulations.
- 9 Cynthia G. Brown, “From ESEA to ESSA: Progress or Regress?” In Hess and Eden, eds., *The Every Student Succeeds Act*, 153–70, 164.
- 10 Chad Adelman et al., *An Independent Review of ESSA State Plans* (Boston: Bellwether Education Partners, 2017).
- 11 Quoted in Alliance for Excellent Education, “New Analyses Identify Shortcomings and Strengths in State ESSA Plans,” press release, June 22, 2017, <https://all4ed.org/press/new-analyses-identify-shortcomings-and-strengths-in-state-essa-plans>.

12 Brandon L. Wright and Michael J. Petrilli, *Rating the Ratings: An Analysis of the 51 ESSA Accountability Plans* (Washington, DC: Thomas B. Fordham Institute, 2017).

13 None of the raters presented a simple “Does this plan do X?” table, although, from our reading, they did provide consistent ratings across states.

14 Correlation coefficients between rating scores are as follows, where 1.00 or –1.00 represents perfect positive or negative correlations, respectively:

	Bellwether	A4ED
A4ED	0.30	—
Fordham	0.22	0.19

15 To create this metric, we calculated a cumulative percent rank for each state in each group’s index. For example, the top-ranked state receives a 1.00, signifying that 100 percent of states scored at or below this state. For our own index, we took the mean value of the three calculations for each state.

16 Alyson Klein, “Arne Duncan Revokes Washington State’s NCLB Waiver.” *Education Week—Politics K-12*, April 24, 2014, http://blogs.edweek.org/edweek/campaign-k-12/2014/04/washington_state_loses_waiver_.html.

17 Adelman et al., “Washington,” in *An Independent Review*, 7.

18 Brian Gutman, “New Concerns in Michigan’s Fourth Take at an ESSA Plan.” *Education Trust—Midwest*, August 25, 2017, <https://midwest.edtrust.org/2017/08/25/new-concerns-michigans-fourth-take-essa-plan>.

19 Quoted in Andrew Ujifusa, “Betsy DeVos Approves Michigan’s ESSA Plan,” *Education Week—Politics K-12*, November 28, 2017, http://blogs.edweek.org/edweek/campaign-k-12/2017/11/betsy_devos_approves_michigan_essa_plan.html.

20 We create the index by creating a standardized score (z-score) for each of fourth-grade reading, fourth-grade math, eighth-grade reading, and eighth-grade math, and then sum the standardized scores for each state.

21 We use the same approach to create the achievement gaps index that we do for the overall NAEP performance index, creating a standardized (z-score) gap score for fourth- and eighth-grade reading and math, and then summing those scores to create a combined index.

22 Achieve, *Closing the Expectations Gap: 2013 Annual Report on the Alignment of State K-12 Policies and Practice with the Demands of College and Careers* (Washington, DC: Achieve, 2013).

23 As part of a 2009 economic stimulus package, the Obama administration received education funds to distribute to the states in a grant competition known as the Race to the Top. To distribute the funds, the Department of Education empaneled stakeholders and officials to rate state grant submissions. In doing so, they gave each state a composite score.

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

25 We constructed our measure by standardizing the results within each individual component of the index, calculating z-scores, then summing the z-scores across all groups within each state to reach a total composite score.

26 Kenneth J. Meier and Laurence J. O’Toole Jr., *Bureaucracy in a Democratic State: A Governance Perspective* (Baltimore: Johns Hopkins University Press, 2006).

27 Data on state party control came from the National Conference of State Legislatures (as of March 31, 2020), www.ncsl.org/research/about-state-legislatures/partisan-composition.aspx.

28 Data on union membership come from Barry Hirsch and David Macpherson, “Union Membership and Coverage Database from the CPS,” (as of February 14, 2020), <http://unionstats.com>.



- 29 Arnold Shober, *Splintered Accountability: State Governance and Education Reform* (Albany: State University of New York Press, 2010).
- 30 S. D. Lawrence, “Iowa’s NCLB Waiver Application: Denied,” *Education News*, June 26, 2012, www.educationnews.org/education-policy-and-politics/iowas-nclb-waiver-application-denied.
- 31 Iowa Department of Education, “ESSA Plan Development | Iowa Department of Education,” 2017. <https://educateiowa.gov/pk-12/every-student-succeeds-act/essa-plan-development> (accessed February 25, 2020).
- 32 Bellwether Education Partners, *An Independent Review of ESSA State Plans: Iowa* (Sudbury, MA: Bellwether Education Partners, 2017), 2, https://bellwethereducation.org/sites/default/files/Bellwether_ESSA_PlanReview_IA_Final.pdf.
- 33 Iowa Department of Education, *Every Student Succeeds Act in Iowa: Final Plan* (Des Moines: Iowa Department of Education, 2018), ix.
- 34 Council of Chief State School Officers, *Equity and Personalized Learning: A Research Review* (Washington, DC: Council of Chief State School Officers, 2017).
- 35 Kim Burgess, “U.S. Education Department Approves NM ESSA Plan,” *Albuquerque Journal*, August 9, 2017. www.abqjournal.com/1045694/us-education-department-approves-nm-essa-plan.html.
- 36 Bellwether Education Partners, *An Independent Review of ESSA State Plans: New Mexico* (Sudbury, MA: Bellwether Education Partners, 2017), 2. https://bellwethereducation.org/sites/default/files/Bellwether_ESSA_PlanReview_NM_Final.pdf (accessed May 22, 2020).
- 37 New Mexico Department of Education, ESSA in New Mexico, Section 4 ESSA Amendments, “Executive Summary of Proposed Changes to Section 4: Accountability, Support, and Improvement to Schools,” 2019, updated February 5, 2020, 1, <https://webnew.ped.state.nm.us/information/essa-new-mexico>.
- 38 Quoted in Kim Burgess, “Reviewers: New Mexico Education Plan Best in the Nation,” *Albuquerque Journal*, June 27, 2017. www.abqjournal.com/1024507/independent-reviewers-new-mexico-education-plan-best-in-the-nation-2.html.
- 39 Alyson Klein, “What Happens If a State Makes Changes to Its ESSA Plan without Betsy DeVos’ Sign-Off?” *Education Week—Politics K-12*, March 27, 2019. <http://blogs.edweek.org/edweek/campaign-k-12/2019/03/essa-new-mexico-devos-changes-plan.html>.
- 40 Klein, “What Happens If a State Makes Changes to Its ESSA Plan”; Shelby Perea, “Feds Give NM the OK on Nixing School Grades,” *Albuquerque Journal*, July 8, 2019. www.abqjournal.com/1337773/feds-give-nm-the-ok-on-nixing-school-grades.html.
- 41 Sunil Joy, *An Analysis of Michigan’s ESSA Plan* (Royal Oak, MI: The Education Trust—Midwest, 2017), 1. <https://midwest.edtrust.org/essa> (accessed February 27, 2020).
- 42 Andrew Ujifusa, “Withdraw and Rework State’s ESSA Plan, Michigan Lt. Gov. Urges,” *Education Week—Politics K-12*, April 28, 2017. http://blogs.edweek.org/edweek/campaign-k-12/2017/04/michigan_withdraw_essa_plan_lieutenant_governor.html.
- 43 Michigan Department of Education, “Michigan Plan to Move Schools Forward Is Approved by the U.S. Secretary of Education.” Press release, November 29, 2017. www.michigan.gov/mde/0,4615,7-140-37818_34785-453658-,00.html.
- 44 Alyson Klein, “Trump Ed. Dept. to Michigan: Your ESSA Plan is Missing Major Details.” *Education Week—Politics K-12*, August 8, 2017, http://blogs.edweek.org/edweek/campaign-k-12/2017/08/trump_ed_dept_to_michigan_your.html.
- 45 Klein, “Trump Ed. Dept. to Michigan: Your ESSA Plan is Missing Major Details.”
- 46 Michigan Department of Education, “Michigan Plan to Move Schools Forward,” 1.

- 47 Bellwether Education Partners, *An Independent Review of ESSA State Plans: Michigan* (Sudbury, MA: Bellwether Education Partners), 4–5, https://bellwethereducation.org/sites/default/files/Bellwether_ESSA_PlanReview_MI_Final.pdf (accessed May 23, 2020).
- 48 Quoted in Ujifusa, “Betsy DeVos Approves Michigan’s ESSA Plan.”
- 49 Brian McVicar, “A–F School Grading Plan a ‘Middle of the Road’ Approach to Accountability, Lawmaker Says,” *Detroit News*, February 8, 2018, www.mlive.com/news/2018/02/a-f_school_grading_plan_a_midd.html.
- 50 Jason Botel, “Letter to Michael Kirst, President of the California State Board of Education, and Tom Torlakson, Superintendent of Public Instruction,” December 21, 2017, 4–5, www2.ed.gov/admins/lead/account/stateplan17/cainterimfeedbackletter.pdf.
- 51 Louis Freedberg, “Q&A with Michael Kirst: Single Rating for Schools ‘Just Too Simplistic.’” *EdSource*, September 6, 2019, <https://edsources.org/2019/qa-with-michael-kirst-single-rating-for-schools-just-too-simplistic/617064>.
- 52 Quoted in Alyson Klein, “Betsy DeVos OKs ESSA Plans for California, Utah,” *Education Week—Politics K-12*, July 12, 2018. http://blogs.edweek.org/edweek/campaign-k-12/2018/07/betsy_devos_approves_essa_plans_california_utah.html?cmp=SOC-SHR-FB.
- 53 Botel, “Letter to Michael Kirst,” 10.
- 54 Botel, “Letter to Michael Kirst,” 3.
- 55 California Department of Education, “California Wins Federal Approval for Every Student Succeeds Act State Plan,” press release #18-48, July 12, 2018, 1, www.cde.ca.gov/nr/ne/yr18/yr18rel48.asp.
- 56 Florida Department of Education, Every Student Succeeds Act (ESSA) State Plan. September 24, 2018, 11, <http://www.fldoe.org/core/fileparse.php/14196/urlt/FL-ESSA-StatePlan.pdf>.
- 57 Bellwether Education Partners, *An Independent Review of ESSA State Plans: Florida* (Sudbury, MA: Bellwether Education Partners, 2017), 2. https://bellwethereducation.org/sites/default/files/Bellwether_ESSA_PlanReview_FL_Final.pdf (accessed May 23, 2020).
- 58 Daarel Burnette, “Florida to Seek Waiver from Key ESSA Provisions,” *Education Week*, July 20, 2017. www.edweek.org/ew/articles/2017/07/20/florida-to-seek-waiver-from-key-essa.html.
- 59 Burnette, “Florida to Seek Waiver.”
- 60 Quoted in Daarel Burnette, “As Florida Submits Fourth ESSA Draft, Governor Urges Fair Treatment from Feds,” *Education Week—State EdWatch*, September 12, 2018. http://blogs.edweek.org/edweek/state_edwatch/2018/09/with_fourth_essa_draft_fl_gov_rick_scott_urges_fair_treatment_from_feds.html.

The views expressed in this publication are entirely those of the authors and do not necessarily reflect the views of the staff, officers, or Board of Overseers of the Hoover Institution.

Copyright © 2020 by Paul Manna and Arnold Shober. All rights reserved.

26 25 24 23 22 21 20 7 6 5 4 3 2 1



About the Authors



PAUL MANNA

Paul Manna is the Isabelle and Jerome E. Hyman Distinguished University Professor of Government at William & Mary. After earning his BA in political science from Northwestern University, Manna taught social studies and coached debate at his hometown public high school in Michigan for three years before earning his MA and PhD in political science from the University of Wisconsin.



ARNOLD SHOBER

Arnold Shober (PhD, University of Wisconsin–Madison) is a professor of government at Lawrence University, where he offers courses on public policy, campaigns and elections, mass media, and research methods. His interests include American federalism, education policy, and federal regulation. His most recent book is *In Common No More: The Politics of the Common Core State Standards* (2016).

Hoover Education Success Initiative

With passage in 2015 of the Every Student Succeeds Act (ESSA), states are again in charge of American education policy. To support them in this undertaking, the Hoover Education Success Initiative (HESI), launched in 2019, seeks to provide state education leaders with policy recommendations that are based upon sound research and analysis.

HESI hosts workshops and policy symposia on high-impact areas related to the improvement and reinvention of the US education system. The findings and recommendations in each area are outlined in concise topical papers.

The leadership team at HESI engages with its Practitioner Council, composed of national policy leaders, and with interested state government leaders. HESI's ultimate goal is to contribute to the ongoing transformation of the nation's education landscape and to improve outcomes for our nation's children.

For more information about the Hoover Education Success Initiative, visit us online at hoover.org/hesi.