

# Designing the Next Generation of State Education Accountability Systems:

## Results of a Working Meeting

September 2014

### INTRODUCTION

#### A NATION'S ACCOUNTABILITY SYSTEMS AT RISK

*“Things have gone so far that having legitimate, real, honest, transparent, and accurate school accountability has become very, very hard.”*

Twenty-five years after the National Governors' Association held a summit on accountability in Charlottesville, Virginia, the systems put in place following that milestone moment in the accountability movement are now at a crossroads.

Skepticism and political pushback have emerged from the left and the right as state accountability systems, school turnaround strategies, teacher evaluation, the Common Core, and standardized assessments have been conflated by skeptics into one overwhelming, unwieldy system with limited results and a host of unintended consequences.

Whether critics' perceptions are right or wrong, growing criticism—including an emerging “mom and dad revolt”—may undermine and ultimately upend the entire concept of state-based accountability unless responsible changes are made, and made quickly. As one policy expert lamented, “We could lose this thing.”

The stakes are high. Despite their limitations, first-generation accountability systems have had a significant impact on schools over the past quarter-century:

- They have contributed to improved academic outcomes for all students, including those who have traditionally been underserved—and, at times, ignored.
- They have provided unprecedented school-, teacher-, and student-level data that has been brought to bear on interventions in the schools with the highest needs, and broader instructional improvement elsewhere.
- They have led to a growing understanding that teachers are key levers of school improvement, and to a newfound emphasis on providing professional learning opportunities to improve instruction.

While much work remains to be done in all of these areas, the need for accountability is only increasing. As schools face unprecedented changes, more and more families are being given the option of school choice, and dramatic shifts in the delivery of instruction are emerging that will transform what schools look like and how they teach students, accountability systems will play a critical role in ensuring that these changes live up to their promise of improving students' lives and giving them a better chance of succeeding in college and careers.

As an important first step, the Center on Reinventing Public Education and the Thomas B. Fordham Institute brought together some of the brightest minds in the accountability world to consider what's needed in the next generation of accountability systems. Many participants played key roles in the creation of existing systems, and none hesitated to identify their shortcomings. Over two days, these experts focused on fundamental beliefs that may help surface design principles for these next-generation systems, as well as the key tensions and challenges that must be addressed for them to become a reality. This report shares current-day challenges, key principles that these next-generations systems should have, and key tensions that must be resolved to ensure that they live up to their promise. Disagreements arose, but on one point there was no disagreement—the time to change systems is now, and the opportunity to do so may not be here forever. We hope that this report will help policymakers and others begin this crucial work.

Washington, D.C.  
June 10-11, 2014

# LIMITATIONS, CONTRADICTIONS, AND UNINTENDED CONSEQUENCES

## THE CHALLENGES FACING CURRENT ACCOUNTABILITY SYSTEMS

*“We built a system we can’t deliver against, and we have to address that.”*

While many of the aspirations of existing accountability systems—particularly those focusing on improved outcomes, equity, and actionable information for school improvement—remain as crucial today as they were a quarter-century ago, we now have the benefit of hindsight to help improve and shape these systems going forward. In order to frame the parameters of a next-generation accountability system, symposium participants addressed the limitations, challenges, and unintended consequences of current state accountability systems. Key contradictions quickly emerged, including:

- State accountability systems serve a range of different purposes, which can conflict with or limit each other’s impact.
- New expectations and assessments have increased burdens on students and schools, without providing enough actionable information in return.
- More sophisticated data collection, analysis, and reporting can yield new insights, but without transparency, it can also foster skepticism.
- States were given broad new powers to intervene in struggling schools, but their capacity to do so in effective ways often remains limited.
- Rigid requirements, when combined with higher expectations, can stifle innovation and lead to efforts to game the system rather than foster systemic change.

Among the fundamental challenges faced by current-day accountability systems:

**Multiple and overlapping purposes** | Accountability systems are expected to drive both educational improvement and public information, even though it may not be possible for the same metrics and measures to do both equally well. “A lot of the complexity comes from the collision of these two ideas,” one participant said.

For example, complex or wide-ranging metrics—such as value-added analyses or a dashboard with multiple indicators—may not serve parents and the broader public well, as they can be insufficiently transparent or overwhelming. But that’s precisely the kind of granular information needed by schools to drive improvement.

Also, the same standardized assessments or other metrics are often used to measure performance of schools, teachers, and individual students. “We should try to do all three, but not with one assessment,” one participant said. However, addressing this issue may conflict with another key issue with current systems—the perception of excessive burdens.

**Excessive burdens** | Test fatigue is one of the key drivers of public dissatisfaction with accountability. Pushback against imposing or changing existing standardized assessments also makes addressing the limitations of existing metrics more challenging.

**Limited value in driving improvement** | Teachers don’t consider annual standardized tests helpful in improving instruction, and parents may draw the wrong conclusions from broad-based measures or efforts to consolidate multiple metrics into a single letter grade. In some places, limited information has led to overlapping accountability systems administered by the state and the LEA—leading to bad optics and confusing results.

**Excessive complexity** | While granular data and other information are needed to drive instructional improvement, some systems are overly complex—one large LEA’s performance framework has 80 indicators, as well as value-added metrics. Excessive complexity may reduce the efficacy of evaluations, make it difficult to provide effective feedback to drive specific improvements, and result in measures that teachers and parents don’t understand, such as value-added/regression analyses that foster skepticism.

**Concerns with state capacity** | While states have been given opportunities in recent years to seek waivers on NCLB requirements and develop alternative metrics, symposium participants cited a “lack of creativity” and an unwillingness to go beyond basic metrics on many states’ part. Many cited these examples as evidence that states never built sufficient capacity to support accountability systems and the implicit expectations that they would provide good data and effective interventions. As one participant said, “When ideas like this hit a low-capacity system, we get compliance—we don’t get improvement.”

**Unintended consequences** | Participants identified three examples of the unintended consequences created by current systems:

- Accountability systems with consequences have led to changed behaviors, as intended. However, depending on the signals that the system sends, those changes can lead to poor instructional decisions. As one participant asked, “What are the signals you want to send to a classroom teacher on what you want to work on with kids? If you’re not clear on that, you’re going to end up with a pretty deep set of countervailing forces.”

- Lack of autonomy forced by consequences can also drive high-performing teachers away from the schools that need them the most, and stifle overall innovation. “Being overly directive will lead to teachers who want to follow steps and mandates, not innovate,” one participant said. “It cascades down and looks like a lack of creativity from here, but it’s a direct response to the regulatory environment.”
- Interventions such as the ones mandated by NCLB have often led to a short-term focus on bumping “cusp” students to proficiency levels, which does not address the needs of the students who need the most help and does not contribute to meaningful systemic change over time.

**Challenges with change management** | Top-down accountability systems may not be the appropriate way to encourage innovation and ground-level improvements. “We’ve tried to do too much via accountability—it encroaches into school change and systems improvement, and there’s a sense out there that there’s too much centralization,” one participant said.

Together, these challenges present an even larger dilemma for states evaluating their accountability systems, participants said. To ensure that accountability systems fulfill their responsibilities, states will have to significantly increase capacity—or scale back the scope of what they directly do. As one participant said:

*“Are we willing to stand by things we can’t effectively execute on because they’re good things? We need to either invest heavily to get it right, or be much more modest.”*

# AN EMERGING FRAMEWORK FOR CHANGE

## DEFINING THE KEY OBJECTIVES OF A NEXT-GENERATION ACCOUNTABILITY SYSTEM

*“We cannot design a perfect system. We can do one that improves over time.”*

For reasons both political and practical, present-day accountability systems have reached a tipping point that imperils the ability of states to continue their historical oversight role in ways that can improve schools and outcomes for students. Even the strongest supporters of accountability—and many of its key architects—now agree that sweeping changes must be made to preserve it. Key elements of what a next-generation accountability system should look like began to emerge during the meeting. There was broad consensus that such systems should:

- *Be built around the child and his or her family*, with the key goal of the system ensuring that each child has the ability to be successful at college and career, and that each child’s parents have sufficient information about the performance of the school to make informed decisions about their child’s education.
- *Continue to support efforts to improve equity*. Systems must continue to ensure that schools address the needs of all children, regardless of race, background, or ability.
- *Emphasize clearly defined objectives for schools and opportunities for students*. Rather than rely on simplistic measurements and automatic, one-size-fits-all punishments, next-generation systems must provide motivation and mechanisms for schools to improve—and opportunities for students to find meaningful alternatives to schools that continue to fail to meet their needs.
- Reflect the idea that *accountability is a core government function*—a responsibility to ensure that public dollars and the welfare of children are being addressed appropriately. Research by Stanford’s Eric A. Hanushek confirms that systems with assessments, standards, and consequences have fared better than those with only some of those components, so it’s critical that next-generation systems focus on all three of these components, albeit in new ways.
- *Emphasize information that drives student choice and improvements in instruction and teacher capacity*. When done well, accountability systems provide the kind of information needed for families to make decisions about their children’s education, for principals to lead their staffs in continuous improvement, and for teachers to improve instruction. In such cases, accountability isn’t a burden on students and schools, but actively helps them make decisions that improve outcomes.
- *Be transparent, fair, yield valid information, and be reliable*. Again, a key goal should be prioritizing transparency and information for those who need it the most—parents, teachers, principals, and the district.

- Not serve as an end to itself, but as a *means to drive motivation and learning*. Next-generation systems will recognize that teachers are the strongest in-school levers to improved outcomes by providing data and resources to improve instructional practice, curriculum, and allocation of resources to serve all students, including those who have traditionally been underserved.
- Are backed up by *sufficient infrastructure to support struggling schools and create options for children in those schools that do not improve*. Either directly or through local systems, next-generation systems must build on their predecessors’ efforts to provide tiered interventions and support for low-performing schools. They must also retain meaningful consequences for schools that fail to meet the objectives of such interventions, including student choice and, in extreme cases, closure.

### A ‘LIGHTER TOUCH?’

There was also an emerging consensus about state accountability systems providing a light (or lighter) touch on districts and schools, without ceding the idea that states should provide oversight and develop supports and strategies to help schools improve. As one participant said, “We need to be humble about what we can do with accountability, but we still have a whole lot of other strategies.”

However, a light touch does not simply mean scaling back existing accountability systems. Instead, participants agreed that a refined approach to accountability that builds on the principles of existing accountability systems but refocuses and narrows their role must be at the core of systems change. These next-generation systems would include:

**School-centered accountability** | Consensus emerged that at the state level, accountability systems should focus squarely on schools—not individual teachers or students. In such a system, responsibility for improvement—as well as broad autonomy—falls on the principal. “We want to create solutions where [building] leaders have an incentive to problem-solve,” one participant said.

In such a system, the state role would focus largely on providing extensive data and creating local agency and flexibility to use it, particularly at the building level. “We’re not giving up—the data will be there, it will be collected and provided to the principal, and then he or she should have the pressure of accountability to do something with that.”

### **Delegating stakes for teachers to the school level |**

While accountability systems have led to a much greater ability to measure teacher effectiveness, many argued that adding stakes for teachers at the state level was an overreach with heavy practical and political consequences. Many participants argued in favor of eliminating automatic stakes for teachers at the state level, instead using the information from accountability systems to provide principals with actionable information about teachers—and the latitude to make human capital decisions at the building level.

### **School choice as a key element of accountability |**

Most participants agreed that states providing information to parents about school performance should also provide options that allow them to act on that information. Research has suggested that the presence of school choice impacts the actions of both parents and school leaders, making it an effective consequence. “Of all the things in NCLB I regret, the biggest one is... the meek little choice provision that we didn’t get and should have insisted upon,” one participant said. “We have got to get real about this.”

### **The continued collection of data—and the use of multiple measures |**

While ceding stakes for individual teachers to the local level, next-generation accountability systems should still gather more—and better—information about schools, teachers, and students that can be used by school leaders to drive improvement. Among possible design ideas:

- Assessments should be domain- or content-based wherever possible, and an emphasis should be placed on their continual improvement.
- Longitudinal systems can give states the opportunity to look beyond test scores and toward outcomes, as well as provide measures of student progress over multiple years that are a more accurate and meaningful way to measure school performance.
- Other options could include other domains of socio-emotional learning and competency-based assessments of mastery, career preparation, dropout recovery, and more subjective measures, such as qualitative reviews.

### **Frameworks and assessments that reflect intentional instructional goals and drive people toward them |**

Participants acknowledged that assessments will always drive curriculum and instructional decisions, so states should focus on ensuring that their systems are focused on sound instructional goals. As an example, the new Common Core assessments were cited as potentially prompting deeper work than typical test prep. As one participant said, “When people try to game it, they get kids to write a lot of research papers.”

### **Consequences that focus on long-term solutions, not only immediate fixes |**

One potential model involves the creation of contractual or plan-based approaches in which individual schools have broad autonomy to specify goals over a multi-year period, with those plans reviewed, approved, and then monitored on an ongoing basis. If those schools cannot organize themselves to perform this work, or fail to meet goals under such a model, consequences would still include closings or other turnaround measures, but such systems would encourage better, more innovative efforts at the local level. “To have an accountability system that actively makes you want to avoid taking risks is a danger in itself,” one participant said.

**Flexibility |** Accountability systems must be able to be revised based on feedback and impact on the ground. That flexibility must also extend to performance metrics, which should be adjusted as circumstances change, rather than establishing fixed levels of performance. To do these things, states must adopt a continuous improvement mindset.

In an era in which different types of public schools have proliferated, accountability systems must also be flexible enough to accommodate different school contexts, including charter schools and instructional models that focus on mastery, accelerated instruction, and personalized learning.

**Future-proofing |** As instructional designs continue to evolve and charters and other schools embrace new organizational models, accountability systems must be able to accommodate innovative modes of instruction and school structures.

First-generation accountability systems have lasted a quarter-century, during which time significant shifts in our understanding of how children learn have accompanied equally dramatic changes in technology, the economy, and the skills and knowledge needed for students to be successful in college and careers. Next-generation accountability systems cannot focus only on what schools look like now, but what they may become as the world continues to change.

## RESOLVING KEY ISSUES

### KEY DESIGN PRINCIPLES NEEDED TO DEVELOP EFFECTIVE SYSTEMS REMAIN THE FOCUS OF SIGNIFICANT DISAGREEMENT

*“We have a moment in time where people in the system are suddenly saying ‘help’... If we can use accountability to change capacity in the system, that could be part of the ticket out of this.”*

If there is a time to suggest sweeping changes to accountability systems, it may be now, as the Common Core and the upcoming round of ESEA waivers provide unprecedented opportunities for states to reframe accountability around providing support to their schools.

These immediate needs mirror a broader, long-term vision of an effective next-generation accountability system in which states provide a formal structure for a far richer decisionmaking process, often led by those who know each school—and the capacity of its leaders to bring about change—best.

During the meeting, some immediate recommendations surfaced—building state infrastructure, making investments in instructionally sound tests and other measures, emphasizing transparency, especially at the local level, and educating the public about the importance of accountability in an era of sweeping change. However, for the promise of next-generation accountability systems to become reality, stakeholders will have to grapple with a range of critical tensions—and do so sooner rather than later. Among the tensions identified:

**The scope of systems and consequences** | Does a leaner, more locally focused approach to accountability imply a state-level emphasis on just the lowest-performing schools, or on all schools? Can the same system fairly assess and impose sanctions on all schools—both low- and high-performing?

Some participants argued that a more streamlined state design would focus consequences and resources almost exclusively on the bottom-tier schools, with the vast majority of schools receiving incentives to implement continuous improvement measures. “We can’t walk away from the bottom 5 percent, but we need a theory of action about how transparency drives improvements in the other 95 percent of schools,” one participant said.

Others expressed concern that focusing narrowly on the bottom 5 percent could allow higher-performing schools to abandon efforts to improve teacher quality. High-performing schools could also lose incentives to actively address the academic performance of minorities, poor students, and other subgroups, in the same fashion as some schools did in the era before NCLB and disaggregation—an unacceptable consequence of a more narrow approach that contradicts a quarter-century focus on equity.

Potential approaches to addressing all schools could focus on those with limited growth, whether they are high- or low-performing, or requiring all schools to report on the progress of their bottom 20 percent of students, whether or not they fall into an NCLB subgroup. Setting targeted goals, such as tracking students’ probability of college admission over time, as compared to other schools that serve similar subgroups of students, is another option.

**What local accountability looks like in practice** | If states delegate most stakes to the school level, does the burden of accountability fall solely on the principal? And what is the role and responsibility of the district, which in most cases has greater control than a building leader over resource allocation and whose decisions about curriculum have a large impact on student outcomes?

As with SEAs, it’s not safe to assume that districts will have the capacity or the will to intervene effectively in struggling schools—particularly small and rural districts, some of which represent individual schools. “It’s a little flip to throw out moving from the state to the district and saying it’s a better logic,” one participant said. “We need to talk about what this means in the real world.” Questions also remain about the accountability of district leaders if they fail to dismiss ineffective principals or provide sufficient oversight.

**Teacher stakes** | If decisions about teacher performance are delegated to local leaders, will those leaders have the same ability to make the kinds of human capital decisions that state-based teacher effectiveness systems have provided? This question is particularly critical given the limitations of past efforts to address ineffective teaching. “It never happened, and if we walk away, it never will happen,” one participant said. “If there’s no teacher effectiveness, I’m not sure you’re going to get the gains in school you ought to have,” another agreed.

**The mechanisms of choice** | Debate revolved around the extent to which choice should supplant other sanctions of an accountability system, and for what schools—just the lowest-performing ones, or all schools? Even among supporters, questions arose about the provision of an “exit visa” to students of schools that fail to improve—would they be able to transfer only to other district schools, to other districts, or would the visa be, in effect, a voucher?

**Level of complexity** | The correct balance between systems being purely informational and overly focused on consequences was a topic of debate. Navigating the tension between the extensive data needed to improve instruction and the goal of transparency with the public also led to considerable discussion. Should all data, such as VAM information for individual teachers, be released to the public—or even to the state in terms of its own evaluations? Some argued that systems—or at least the public-facing parts of them—need to be simple enough for parents and teachers to understand, lest they foster skepticism and mistrust. Others argued that transparency trumps other factors, with a high burden of proof for not making information public.

Complexity also factors into how consequences are administered, with district leaders equally eager to push for exceptions and to criticize systems as being overly complex.

**Goal setting** | particularly considering the impact of the Common Core and other more rigorous standards. “We have to be very careful about how to introduce accountability to new higher standards,” one participant said.

How can an accountability system set a higher bar without imposing blanket consequences for those who fail to meet them? “We confuse aspirational goals—all kids will be career and college ready—with achievable goals,” one participant said.

Consensus emerged that goals in an accountability system should be realistic, but still incorporate challenging “stretch” goals with a differentiated system of rewards and penalties for aspirational and achievable goals. Some suggested that school performance should be criterion-referenced, or measured against the best performance of schools with similar characteristics. Others advocated for a dual “carrot and stick” system that emphasizes rewards for higher levels of performance and limits consequences to failure to meet basic standards. And one observer noted that the Common Core’s focus on progressions could actually provide a new, more realistic set of targets.

**Appropriate uses of measurements and metrics** | Concepts like “progress,” “velocity,” and “growth” can conflict with more static measures of achievement. “I’m deeply suspicious of growth measures that don’t end up at the point of college readiness,” one participant said. At the same time, current systems and metrics (including the 100 percent goal of NCLB) do not provide the public with an accurate picture of how far behind many children are. As one participant reflected, “It’s a bad choice—you go with something politically survivable, or you go with something that meets kids’ hopes and dreams?”

Short-term growth measures also don’t always measure absolute performance—in part because of limitations on what tests can capture. Even in high-performing districts where low-income children now have higher learning gains than more affluent ones over short periods of time, the overall impact on performance has been limited. Some suggested that a measure of “on-trackedness”—or growth toward standards over multiple years—should be the key, especially given that low-income children often lack the velocity to maintain proficiency over time even if they reach it. Others suggested that brutal honesty, while politically difficult, remains the only way to force systemic change and can be achieved by “calibrating consequences.” And states will have to invest in making sure these metrics are valid and reliable—or minimize their use.

**The impact of efforts to reduce burdens** | While there was support for minimizing burdens on students and schools wherever possible, strong disagreements emerged about ways to do so within the context of an effective system.

Potential solutions included every-other-year testing, testing during key “gatekeeper years,” as other countries do, and sampling of students for assessments. Most suggestions met strong criticism—particularly the idea of sampling, which could potentially yield a statistically valid measure of a district’s performance, but loses the ability to track individual student growth over time, or the efficacy of specific teachers, programs, or initiatives. The approach would also be “a disaster” for public-facing accountability and for individual parents, one speaker said. “The parent has no clue on how a child is doing, or knowing when a child falls off the ledge.”

While the quality of existing assessment instruments remains far from ideal, they remain a necessary part of accountability systems—and must continue to be improved over time. Some attendees pointed out that more complex tests that provide better instructional signals may be more burdensome, but “worth it” because of the shift in instruction they could foster. States will have to carefully weigh the trade-offs between better information and imposing additional burdens on students and schools, but never lose sight of the importance of improving existing assessments.

**Flexibility** | While generally agreed upon, some argued that providing local leaders with greater flexibility to develop improvement plans for struggling schools must be balanced with much stronger oversight and intervention strategies when those schools don’t perform. Autonomy must also be grounded in research-based expectations of what improves instruction. “If the price of autonomy is continued bad practice, it’s not worth it,” one participant said.

## CONCLUSION

### THE TIME TO ACT IS NOW

*“Kids only have one chance. We can’t wait to make some decisions.”*

Accountability systems are complex by nature, and changing them involves issues of capacity, logistics, core beliefs about the role of states, districts, and schools, and practical and political considerations. But that was also the case a quarter-century ago, when these systems were created from scratch at the federal, state, and local levels. It was possible then, and it is possible to remake them now.

Some of the pieces needed for this transition will soon be in place, including new assessments aligned with more rigorous standards, but the fundamental tensions and issues surfaced in this report will not resolve themselves. Policymakers, funders, educators, and others will need to grapple with them now, and make difficult decisions that are capable of accommodating both current needs and future changes—in some cases, dramatic—in how schools are structured and deliver instruction.

Next-generation systems can give us the opportunity to return the focus of accountability to students, families, and the public good. They can move states and districts away from checklist compliance and toward fostering innovative approaches to improving teaching and learning. But to do these things, we must start building systems now, and build them with the flexibility to change as our tools and our needs evolve. As one participant said, “We cannot let the perfect be the enemy of the good.”

## PARTICIPANTS

Charles Barone, Democrats for Education Reform  
Stephen Bowen, Council of Chief State School Officers  
Naomi DeVeaux, DC Public Charter School Board  
Chester E. Finn, Jr., Thomas B. Fordham Institute  
Betheny Gross, Center on Reinventing Public Education,  
University of Washington Bothell  
Jane Hannaway, American Institutes for Research  
Lisa Hansel, Core Knowledge Foundation  
Paul Hill, Center on Reinventing Public Education, University of  
Washington Bothell  
Joe Jones, Center on Reinventing Public Education, University  
of Washington Bothell  
Sandy Kress, Akin Gump Strauss Hauer & Feld LLP  
Robin Lake, Center on Reinventing Public Education, University  
of Washington Bothell  
James Liebman, Columbia Law School  
James Merriman, New York City Charter School Center  
Chris Minnich, Council of Chief State School Officers  
Amber Northern, Thomas B. Fordham Institute  
Harry Patrinos, World Bank  
Scott Pearson, DC Public Charter School Board  
Michael Petrilli, Thomas B. Fordham Institute  
Katie Piehl, National Association of Charter School Authorizers  
Morgan Polikoff, University of Southern California  
Robert Pondiscio, Thomas B. Fordham Institute  
Van Schoales, A+ Denver  
Nelson Smith, National Association of Charter School  
Authorizers  
David Steiner, Hunter College School of Education  
Shael Suransky, Bank Street College of Education  
Joanne Weiss, Weiss Associates  
Richard Wenning, BeFoundation  
Gene Wilhoit, National Center for Innovation in Education,  
University of Kentucky  
Judy Wurtzel, Independent Consultant

## ACKNOWLEDGMENTS

We're grateful to those who took part in the meeting summarized here. Participants brought sharp ideas and thoughtful reflections to the discussion. The meeting was a direct result of a lively email debate prodded by Michael Petrilli, president of the Thomas B. Fordham Institute. The Fordham Institute hosted and helped organize the meeting. The Laura and John Arnold Foundation provided financial support, but the views represented here are not necessarily those of the Foundation. Finally, Mark Toner (with guidance from Joe Jones) wrote this excellent, succinct summary of our complicated, and sometimes meandering, discussion.