DO-OVER OR DOUBLE DOWN?

Working Toward a New K–12 Education Accountability Ecosystem

Michael Q. McShane
Paul DiPerna

edCHOICE

ABOUT EDCHOICE

OCTOBER 2018

EDCHOICE.ORG

DO-OVER OR DOUBLE DOWN?

Working Toward a New K–12 Education Accountability Ecosystem

Michael Q. McShane
Paul DiPerna

ed CHOICE

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

EXECUTIVE SUMMARY

These are uncertain times for school accountability. At the federal level, the Every Student Succeeds Act (ESSA) returned authority and responsibility for school accountability to states, breaking from a multi-decade pattern of increasing federal involvement.

At the same time, the standardized tests that have been the backbone of school accountability systems are losing popularity and public trust. Without reliable metrics upon which to base accountability systems, the whole effort will crumble.

Within states and districts, each passing year offers more educational options for students and families. Families can choose from traditional public schools, magnet or specialty schools, career and technical education-focused schools, public charter schools, private schools, or a host of decentralized education providers.

This evolving world of increased choice and diversity in the options available to families raises numerous serious questions about the future of educational accountability:

- What does it mean to hold schools accountable?

- Why do we even care about this issue? Is it because taxpayers fund public schools? Because we care about what or whether children learn? Because we're worried parents don't know what is going on in their child's school?

- What does accountability look like in schools and education systems where families have more control over their child's learning environment?

- How do we measure school performance?

To wrestle with these difficult issues, we convened two meetings, each with four focus groups of K–12 stakeholders who participated in a series of small and large group discussions facilitated by

Hanover Research. This report is a summary of their experiences, observations, and opinions. The participants fell into four categories, or affinity groups:

Engaged Outsiders — Current non-educators, but work or volunteer in education-related organizations. They tend to have worn multiple professional hats in their established careers. Most are advanced in their careers and have achieved executive-levels positions. Some have launched their own education technology or service-providing companies.

Practitioners — Direct experience in schools, districts, or schooling organizations across various sectors, including public schools districts and district schools; public charter schools; private schools; and blended schools. Some are school leaders, teachers, school board members, or current and former superintendents.

Policy Advocates — Recognized education policy experts, working in educational advocacy organizations, state think tanks, statewide membership associations, or national public policy organizations.

Researchers — Education researchers who have specialized in a variety of fields including accountability in K–12 education. They are based at universities, state or national think tanks, or other nonprofit organizations that conduct education research and analysis and advance thought leadership.

Participants lauded increased focus and attention on educational equity and achievement gaps between different student subgroups, greater transparency, improved data literacy, the proliferation of successful schools serving high-poverty populations, and improved test scores. They also highlighted some missteps and unfortunate trends including the narrowing of the goals of schooling, incentivizing counterproductive behavior like "drill and kill" or gaming, and the ill-executed attempt to hold teachers accountable.

Participants highlighted a lack of clarity of purpose for accountability systems, an unfortunate layering

of local, state, and federal mandates and tests that suck up time and energy, the fact that parents don't seem to care a great deal about standardized test scores, and the political problems that a reform like accountability (that doesn't have a natural constituency) engenders.

Our focus groups arrived at some consensus in their assessments of the past and present of accountability, but they disagreed on what to do going forward. The first convening of focus groups tended to prioritize the outcomes of an accountability system. These participants felt that school accountability should ensure all schools prepare students who are ready to succeed in college, the workforce, and as citizens. By contrast the second convening of groups spent much more time discussing and debating the underlying processes and implementation of accountability. In both convenings, these groups voiced frustration with the current approach to accountability. They agreed there is a disconnect between the goals of policymakers and those of parents, teachers, and school leaders. They also argued that the perception of the current systems is that it reflects a top-down, one-size-fits-all approach.

Some other differences emerged by affinity group type. Practitioners and Engaged Outsiders often focused on the "who" of accountability. Should parent preferences be taken more seriously? Should teachers have more voice in goal-settting? Policy Advocates were more likely to focus on their areas of expertise concerning the policymaking process, implementation, and evaluating outcomes. What outcomes are measured? Are those the outcomes we really care about? Can those outcomes be gamed or used as support for counterproductive activities? Researchers were very focused on matters of design and measurement of outcomes. Are measures aligned to goals? Are measures accurate? Are they restrictive? How much of the total scope of what we want from schools shows up in them?

Focus groups could agree on some general principles and potential paths forward, but they also showed diverse views on what to prioritize and the more specific details to operationalize those principles and values.

To help move accountability policy forward, participants identified new priorities and potential metrics of school performance that could, over time, restore public trust and foster school improvement. Restoring trust is at the top of the to-do list for accountability systems, as is more honesty about the tradeoffs inherent in accountability policy. Accountability proponents need to remember that schools are focusing on forming human beings, and that students are more than test scores. School accountability needs to build on the strengths of America (like entrepreneurship and creativity), serve all of the nation's children (including suburban and rural students), and emphasize continuous improvement over carrots-and-sticks mechanistic accountability.

To measure progress, participants identified numerous potential metrics (listed in full on pages 17–20). These include student and teacher retention, parent satisfaction surveys, course offerings, and much more.

Ultimately, we offer as many questions as answers, including:

- Is the purpose of accountability to establish a minimum standard or to drive schools to be better?

- Should school scores be rolled up into singular grades or ratings, like A-F ratings, star ratings, or numerical ratings?

- Can we have strong accountability without imposing uniformity?

- Do we trust parents to know more than the school system about what is best for their child?

- What should be the unit of analysis? The school district? The school? The learner?

- What should schools have to demonstrate in exchange for taxpayer dollars?

Clearly, we have much left to discuss. Accountability policies, and the debates over those policies, will not go away any time soon.

INTRODUCTION

Across the political spectrum, politicians love to talk about "school accountability." In a 2002 speech, then-President Bush said, "one of the cornerstones of any good school system is accountability." [1] President Obama echoed these comments in 2012, stating that "Standards and accountability—those are the right goals." [2]

The phrase "school accountability" has become so engrained in the language of education that it's easy to forget there was a time before standardized testing, school grades, and state intervention in schools.

At first, authorities simply desired to learn information as to what was going on in schools. At the federal level, a department of education was created by an act of Congress in 1867 "for the purpose of collecting such statistics and facts as shall show the condition and progress of education." By the 1920s, the then-named Office of Education was collecting a breakdown of school spending across the country. In 1954, it launched the first annual survey of school enrollment, teachers, and number of school houses. Access to government-collected data and information about student achievement really took off when the National Assessment for Educational Progress (NAEP) launched in 1969. [3]

As data proliferated, authorities began to consider how such information could be used to manage schools. In 1965, during the debate around the Elementary and Secondary Education Act, Senator Robert Kennedy was quoted as saying "Look, I want to change this bill because it doesn't have any way of measuring those damned educators like you, Frank, and we really ought to have some evaluation in there, and some measurement as to whether any good is happening."[4] In the watershed "A Nation at Risk" report, drafters advised in 1983:

"Standardized tests of achievement (not to be confused with aptitude tests) should be administered at major transition points from one level of schooling to another and particularly from high school to college or work. The purposes of these tests would be to: (a) certify the student's credentials; (b) identify the need for remedial intervention; and (c) identify the opportunity for advanced or accelerated work." [5]

In late 2001, the U.S. House of Representatives passed the No Child Left Behind Act (NCLB) by a 381-41 margin. The U.S. Senate passed it 87-10. President George W. Bush signed it into law in early 2002. NCLB essentially nationalized Texas's accountability system, with testing in reading and math in grades 3-8 and again in 10th grade and a series of consequences for schools and districts that did not make "adequate yearly progress" toward the goal of universal student proficiency in the year 2014.

At the dawn of No Child Left Behind's passage in 2001, states across the country offered a patchwork of accountability systems. Some states, like Texas, offered a program of standardized testing with repercussions for both students and schools. Others, like Minnesota, tested students in a few grades but had no repercussions for schools. Still other states, like Iowa, had no statewide standardized testing or accountability frameworks for schools at all. [6] No Child Left Behind changed the landscape across the states.

The Obama administration's Race to the Top initiative utilized regulatory waivers to NCLB, bringing teachers into the fold as well. Those waivers incentivized states to create evaluation systems to measure teacher performance and hold them accountable for how they were serving students.

FIGURE 1	Key Dates in the History of School Accountability

1865 New York launches Regent Exams for 8th graders (adds high schoolers two years later).

1867 Department of Education Act creates a federal department of education "for the purpose of collecting such statistics and facts as shall show the condition and progress of education in the several States and Territories."

1897 Minnesota enacts law requiring schools to meet minimum requirements to receive state aid.

1954 Federal "Office of Education" conducts first annual survey of enrollment, teachers, and number of school houses in America.

1969 First administration of the National Assessment of Educational Progress (NAEP).

1979 Texas begins to administer the Texas Assessment of Basic Skills in grades 3, 5, and 9, starting the "era of school accountability" in Texas.

1983 "A Nation at Risk" report advocates that "Standardized tests of achievement (not to be confused with aptitude tests) should be administered at major transition points from one level of schooling to another and particularly from high school to college or work."

1989 President George H.W. Bush convenes nation's governors in Charlottesville, VA. to discuss education reform, including burgeoning accountability movements in states.

1993 Massachusetts adopts comprehensive K-12 standards (previously only had standards for history and physical education).

1994 President Clinton signs *Goals 2000* into law, including a provision that "All students will leave grades 4, 8, and 12 having demonstrated competency over challenging subject matter including English, mathematics, science, foreign languages, civics and government, economics, the arts, history, and geography."

1999 Florida becomes first state to give A–F letter grades to schools.

2001 *No Child Left Behind* requires all schools receiving Title I dollars to set goal of 100% proficiency in reading and math by the year 2014, to test all students in these subjects yearly in grades 3-8 and again in high school, to make "adequate yearly progress" toward 100 percent proficiency. Failure to do would lead to a "cascade of remedies" escalating with each year of failure.

2011 The Obama administration begins to offer conditional waivers to sanctions of *No Child Left Behind* to states that agree to adopt "college and career ready" standards, revamp their accountability systems, and work to hold teachers and principals accountable in addition to schools.

2015 The *Every Student Succeeds Act* returns accountability policy to states, requiring annual testing, but granting leeway in how states uses those results to hold schools accountable.

What Is Accountability, Anyway?

A good working definition of school accountability comes from the *Handbook of the Economics of Education*: "The process of evaluating school performance on the basis of student performance measures."[7] Central authorities collect data on how students are performing and then judge schools based on it.

Unfortunately, even that good working definition raises as many questions as it answers: *How do we measure school performance? Which schools are we talking about? Who is doing the evaluating? For what purpose?*

And, perhaps on a deeper level, there is the question of why accountability is necessary in the first place. Advocates for accountability argue that states have constitutional requirements to provide an education to the children residing in them. Families are required to send their children to public schools if they cannot homeschool them or pay for a private education for them. As a result, there is some obligation to make sure that those schools provide a quality education for students. As the nonprofit organization Education Trust—West put it when talking about accountability policy for the state of California:

> "An accountability system is how we uphold the constitutional right of education for all California's students – it provides visibility into how districts and schools are doing, helps educators learn from the expertise of stellar schools, identifies struggling schools, and prompts additional supports for those schools."[8]

A more basic definition of accountability, from the Merriam-Webster dictionary, might help clarify. That dictionary defines being accountable as "subject to giving an account, answerable."[9] Schools are held accountable by having to give an account of what they do and what results they are receiving. They must answer for their actions.

Where We Are Right Now

Four forces are at work shaping the contemporary landscape of school accountability. First, new and varied schooling options are emerging in both the traditional public school sector and in challengers to the traditional public school model. Second, parents increasingly have the option to choose between different providers, changing the relationship between the government and schools. Third, the tests that are used to measure school performance are losing legitimacy and public trust. Finally, the federal government has pushed the decisions that must be made about designing accountability systems within this landscape back to states after more than two decades of direction from Washington. Each is worth discussing in turn.

More and more options are becoming available to students around the country. Just under 3.5 million students attend some 4,300 magnet schools.[10] Over 3.1 million students attend some 6,900 charter schools.[11] Almost 500,000 use some form of private school choice.[12] In some metro areas, like New Orleans, Indianapolis, Miami, and Washington D.C., families can choose, with public support, from all of these options. Many of these schools are organized differently from each other, have different missions, different pedagogical philosophies, and a host of other unique features that make apples-to-apples comparisons challenging. Any accountability system that will try to evaluate all of these school types will almost certainly miss important nuances that set these schools apart from one another.

Rather than being assigned to a specified school option, parents are increasingly deciding where their children go to school. Many in the school choice movement have supported accountability in the public system because parents cannot "vote with their feet."[13] If parents are required to send their children to a particular school, then whatever authority requiring them to do so has some obligation to make sure that the school is good. Now parents decide what they want from schools and choose accordingly. But this raises issues. Insofar as families want those decisions supported by tax

dollars, the body politic might want assurances that the money is being put to good use.

Is a parent's imprimatur of a school enough? On that question there are real disagreements within the education reform community. In recent years tensions have arisen between those advancing choice-based accountability driven by parents and families and those supporting rule-based accountability led by state and local policymakers.

In a system where people want reliable information about schools, either to make informed choices or to hold them accountable, metrics become important. Increasingly, though, the primary metric that schools use, standardized tests, is coming under fire. According to the 2017 *Phi Delta Kappan* poll, only 42 percent of Americans think that performance on standardized tests is an important indicator of school quality. [14] In polling that we at EdChoice have conducted on parent opinions, standardized test scores rank very low on parental priorities when choosing a school. [15]

Amidst all of this, the federal government has returned decision-making power regarding school accountability back to states. In late 2015, President Obama signed the Every Student Succeeds Act into law, which did away with federal requirements for schools making Adequate Yearly Progress toward the goal of 100 percent proficiency as mandated by No Child Left Behind. It also eliminated the cascade of federally determined sanctions on schools that was associated with failure to meet those goals. States are still required to test students in grades 3–8 and again in high school, but what they do with that information is up to them. The law also creates the opportunity for states to pilot new, innovative systems of assessments in a subset of their districts before scaling statewide.

The big questions of accountability (Who should be accountable? Accountable to whom? Accountable for what?) remain unanswered. The tensions between what taxpayers (and politicians) might want from schools and what parents might want from schools remain unresolved. And we are losing

confidence in our measurements, risking a debate that is based in anecdote and hearsay.

With wide latitude in policymaking and great diversity in educational options, now is the time to try and dig into these questions and find both areas of agreement and areas of disagreement. Understanding the fault lines can help policymakers better craft accountability policy that reflects current knowledge and embodies contemporary values.

Our Approach and Methods of Inquiry

EdChoice partnered with Hanover Research to design and facilitate a consecutive two-day focus group series to explore the past, present, and future of accountability in K–12 education. We recruited participants based on similarities in their professional experiences. (Hereafter, we use the terms "focus groups," "small groups," and "affinity groups" interchangeably.) The people in these affinity groups did not typically deal with K–12 accountability public policy matters regularly, if at all. Rather, we sought people who have been known to excel in their professional fields, who have been thoughtful and insightful in their careers, and who would be open to sharing experiences and learning from others' experiences. In short, we wanted to engage with others who have sharp perspectives outside the K–12 accountability policy talk bubble. We recruited four affinity group types: Engaged Outsiders, Practitioners, Policy Advocates, and Researchers.

We share detailed information and descriptions of our methods in Appendices 1, 2, and 3.

TABLE 1 — Focus Group Participants

Affiliations listed for identification purposes only

Participants	Affiliation
Amy Anderson	ReSchool Colorado
Kaitlin Anderson	Michigan State University
Tony Bennett	MGT Consulting
Ashley Berner	Johns Hopkins Institute for Education Policy
Beth Blaufuss	Partnership Schools NYC
Katie Boody	LeanLab Education
Allen Bourff	Hamilton Southeastern Schools
Caitlin Codella	U.S. Chamber of Commerce
Dale Chu	DC Strategies
Angela Dills	Western Carolina University
Catherine Durkin Robinson	Florida Parent Network
Gwen Eaddy Samuel	Connecticut Parents Union
Max Eden	Manhattan Institute
John Elcesser	Indiana Non-Public Education Association
Tillie Elvrum	National Coalition for Public School Options
April Garcia	Alliance for Catholic Education
Kevin Hesla	National Alliance for Public Charter Schools
Collin Hitt	Southern Illinois University School of Medicine
Brian Jodice	Parents for Educational Freedom in North Carolina
Tim Keller	Institute for Justice
Brian Kisida	University of Missouri
Robert Maranto	University of Arkansas
Amy McGrath	Arizona State University
Deborah M. McGriff	NewSchools Venture Fund
Wade Moore	Urban Preparatory Academy
Matthew Nelson	GreatSchools
Andrew Neumann	Open Sky Education
Susan Pendergrass	Show-Me Institute
Anthony Pienta	The Philanthropy Roundtable
Karega Rausch	National Association of Charter School Authorizers
Lizzette C. Gonzalez Reynolds	Foundation for Excellence in Education
Marilyn Rhames	Teachers Who Pray Inc.
Ian Rowe	Public Preparatory Network
Michael Schuttloffel	Council for American Private Education
Juliet Squire	Bellwether Education Partners
CJ Szafir	Wisconsin Institute for Law and Liberty
Vanessa Threatte	SUNY Charter Schools Institute
Jill Turgeon	Loudoun County Public Schools
Brittany Vessely	Catholic Education Partners
Eric Wearne	Holy Spirit College
Betsy Wiley	Institute for Quality Education
Maria Worthen	iNACOL

Note: Five focus group participants chose to remain anonymous for the purpose of publication.

LESSONS LEARNED FROM THE LAST 20 YEARS OF SCHOOL ACCOUNTABILITY

Our small group sessions began by looking backward. What have we learned over the last 20 years of school accountability policy? What were the successes? What were the stumbles?

Successes of the School Accountability Movement

Participants from across the ideological spectrum made clear at the outset that school accountability policy did lead to some improvements in schools and education policy. We'd like to highlight four. [16]

1. Focus and Attention on Equity

Participants formed some consensus on the first and most prominent success of the school accountability movement: collecting student test scores and other outcome measures that shine light on low-performing schools. The attached consequences also pushed those schools and communities to better serve their low-income and minority students.

As one engaged outsider stated:

> "I think that a triumph would be the fact that it's no longer okay to let poor kids of color just do poorly. The numbers are there. People are looking. I think it just raised a level of consciousness about how bad this situation has been—besides just this NAEP gap. You could see more data, and more people have been talking about it. More organizations have cropped up over the years to say, 'There's an equity gap.' Whatever kinda gap it is, be it gender-wise, racial, socioeconomic gap, it has produced more voices to advocate for disenfranchised communities and children."

A practitioner echoed these sentiments:

> "And there's a focus on the gaps. There's a real emphasis on equity. That was probably not there prior to the accountability movement."

A policy advocate put it this way:

> "Because even though people say absolute scores on reading and math should be broadened, at the end of the day, we know that our black, brown and poor kids are not achieving at the same levels in just our very basics, reading and math, as kids from wealthier families."

A researcher argued:

> "I think the information that came out of No Child Left Behind was incredibly valuable—with respect to achievement gaps and test scores."

That said, views about equity were complicated. What does it really mean to have an equitable school system? What happens when families define what they want from schools differently? If a student attends a "high-quality" school that is not a good fit for him or her, how do we think about that problem? What if different families have different goals for their children? How should an accountability system react if these differences seem inequitable?

And at least one policy advocate argued that sometimes by focusing on narrow metrics of school performance we risk missing the forest for the trees:

> "Another thing that's really frustrated me in policy conversations post-NCLB is that the meaning of the word 'equity' has become reduced to grade-level outcomes. And it has allowed us to avoid much harder and more nuanced conversations about the things that actually happen in school, and the structures of institutions in the society that contribute to inequity. We're confusing equity and equality."

2. More Transparency

A second success of the school accountability movement has been greater transparency. Prior to the school accountability movement, participants argued, parents and community members had far less information about what was actually going on in schools.

As one researcher put it:

> "If accountability systems are good for anything, it's good for creating transparency."

An engaged outsider argued:

> "Having that transparency, having people who are excited about trying to tackle this and create new school models, has opened up options for parents that didn't exist before."

Another researcher argued:

> "I think one of the things that was striking to me is that, one, the report cards started coming out. This is clearly information that parents didn't know. People responded to it in a way that indicated that it wasn't known information, how well the schools were doing or how well they were doing with respective particular subgroups in the school."

One policy advocate highlighted:

> "The movement towards transparency and more information. Again, we can say, there needs to be more than test scores. But, I think the accountability, the first part of accountability is getting the information out there. I think that has been a step in the right direction."

In fact, one researcher argued, the only reason that we're able to have a lot of the conversations that we do is because of the data created by school accountability systems:

> "And, you know, we've got all this new stuff coming out about the difference between schools that get high test score achievement and other kinds of attainment. And we only know that because we have all this test score data to work with."

3. Data Literacy

Not only has accountability produced more data on school performance, but it also has encouraged conversations around that data and the use of that data to try and improve school performance.

As one practitioner argued:

> "We've probably raised the level of data literacy, and those conversations probably didn't happen before No Child Left Behind, especially for the subgroups. So there's some measurement and data and data literacy I think that are all positive outcomes."

One researcher did add a word of caution:

> "I think, one of the things that I was noticing in [STATE] now, the ability of parents to use data and their desire to use data, varies so much...I don't think we still fully understand what parents are wanting and how we should present the information to them in the most meaningful way."

Another researcher added:

> "The last portion of our conversation, we started talking about accountability and then we got into how parents are using the data. And I feel like, to me, that really raises up the tension between metrics you use for accountability from the government or the state versus metrics that you use for parent information and transparency. And, to me, those two things are very, very different—but often get conflated."

It will be important moving forward to understand that when data are used in systems that reward or punish schools that data run the risk of being gamed and distorted. Are there ways to get the informational benefits of education without the distortion? Participants varied in their assessment.

4. Improved Test Scores

School accountability systems also were credited with making improvements in student performance.

As one researcher put it:

> " I think if you look at the research literature, I think it's hard to argue against accountability raises test scores and the subjects that it sanctions or that it's applied to. I think we don't, exactly, know why those test scores go up. Are students learning more, are schools better at teaching the test? Probably some combination of both. "

Missteps of the School Accountability Movement

While participants highlighted clear successes of the school accountability movement, they were also quick to list some of its missteps. Three major issues are worth highlighting.

1. Narrowing the Goals of Schooling

Participants repeatedly argued that an unintended negative consequence of the school accountability movement has been a narrowing of the goals of schooling to focus only on maximizing reading and math test scores.

One practitioner argued:

> " The current state accountability system…has a narrowing effect [on] high-quality curriculum and what's ultimately taught. "

One policy advocate argued:

> " We want to test reading and math, because we think those scores are important. But then the perversion comes when schools focus in on reading and math, and lose the arts, and lose parts of the sciences. And then within reading and math, they focus on the way to deliver the best test results within those subjects, which can actually harm comprehension, and harm true literacy with a capital L. "

There is debate as to what the goals of schooling should be. As one policy advocate argued:

> " But when we first think about what is education for, and especially as citizens, it wasn't towards metrics of achievement in reading and math. It was for the purpose of self-government. "

Another policy advocate put it this way:

> " There's tension when we back up a step from that and ask the question of what is the purpose of our education system? Yes, there's the folks who'll say it's the basic reading and math skills, but then there're folks who'll say it's about producing a really engaged citizenry who know about our government and how to participate, rather than folks who say it's about workforce and jobs. "

A researcher argued:

> " I think that when there's debates over accountability, it often has its roots in debates over what people think schools should do. So, disagreements over accountability systems are, often…a much deeper issue over how much they focus on math and reading versus the arts and civics versus extracurricular things. And people have [a] different sense of priorities with respect to that. And those different priorities manifest in accountability systems. Because the NCLB accountability system, basically, privileged those who think test scores and preparation for college and workforce is the primary goal of schools. "

One researcher summarized it this way:

> "Accountability homogenizes schools."

2. Incentivizing "Drill and Kill" and Gaming

Related to the narrowing of the goals of schools, numerous participants argued that the incentives built into accountability systems lead to schools engaging in ultimately unproductive activities—both outright manipulation of metrics and softer forms of "gaming" that, while not disingenuous, are certainly counterproductive.

One researcher argued:

> "I definitely think a lesson learned is that people will game the system any way possible."

One policy advocate argued:

> "Yeah, so graduation data is very easy to manipulate. You can graduate by not holding them to any standards. Suspension data is very easy to manipulate, too, because you can get suspensions down by not suspending students. Test scores are harder to manipulate but still possible to manipulate through softer means."

As one engaged outsider put it:

> "Some of our schools that are doing quite well on accountability, you walk into and they're drill and kill factories. That's not the educational experience...we're trying to get at, but on paper they're doing quite well."

One practitioner detailed how evaluation systems can be created that prevent real consequences from trickling down to anyone:

> "Well, I can tell you one of the things that happens. Evaluation systems are created with built-in buffers to keep that accountability from reaching to the classroom level or to the individual level. Rubrics are developed that can sometimes measure inputs rather than the outputs. You can minimize the student performance, in some cases, with these evaluation instruments. So, some very ingenious evaluation tools have been designed."

3. Expanding Accountability to Teachers Drove Pushback

Another misstep that participants brought up repeatedly was the Race to the Top-era decision to include teachers in the accountability framework. Pushing for accountability at the school level is one thing; doing so at the teacher level is another entirely.

One engaged outsider put it this way:

> "I think the perception of the nationalization of standards and assessments and tying that to teacher evaluations was a huge, huge mistake."

A policy advocate argued:

> "If a student's performance, academic performance, had no bearing on a teacher's pay or teacher's performance or teacher's ranking, no one would care about accountability or assessment. They'd be all for it. Teachers would say "these are useful tools, I need to know where [NAME] is and isn't, I need to know how to help her more." So I think we took a wrong turn in that major connection between accountability for student achievement and attaching it to the adults, the teachers. And then it just kind of got lost, because people deal in self-interest."

Another policy advocate built on those sentiments, stating:

> "You saw that reaching, rising and rising with the Bush administration, only accelerated in the Obama administration when it gets to the point of, 'Okay, well it hasn't worked when we've tried this on schools. Now let's try this on teachers.'"

Talking about trying to attract high quality teachers, one researcher argued:

> "The pipeline has declined dramatically in states that have taken these sorts of teacher evaluation things seriously."

Another researcher argued:

> "I think a lot of policy folks think teacher evaluation to be a separate thing from accountability, but I think from a teacher's standpoint, it's all wrapped up together. A lot of states have moved away from teacher evaluation, to some degree, but they still are fearful...I think it's one of those things that will take a while to walk back, if that's the route we want to go now."

4. A Mixed Bag: Trust

Opinions differed on one lesson from accountability policy thus far, and that was the issue of trust. On one hand, participants argued that accountability policy was a trust-building exercise, particularly with low-income communities. It told them that their children mattered and would not be left behind. On the other hand, missteps and challenges in accountability policy eroded trust in central authorities to accurately measure school performance and act on it.

One policy advocate put it this way:

> "I believe what happened with 'No Child Left Behind,' the term...was intentional: 'No Child Left Behind,' black and brown children are going to get a chance. I took that term very literally. As I understood, was it was meant to make sure that children that were once forgotten in the classroom weren't forgotten anymore because you had to have accounted for every person in the seat."

One researcher argued for the other side of the coin:

> "I feel like one lesson from the accountability movement is that policymakers and technocrats from state capitals in D.C. have lost credibility. Because they designed these systems that people don't believe in and insisted that they're the right way to do things and given grades and all of this stuff. So, I just think one lesson learned is humility. We've gotta be a little bit more humble from our perch in D.C. or other places."

THE CURRENT STATE OF SCHOOL ACCOUNTABILITY

Small group participants highlighted four issues that are central to current debates about school accountability. They argued that accountability policy lacks a necessary clarity of purpose. They bemoaned the complex and sometimes competing mandates of localities, states, and the federal government. They argued that there is a disconnect between what parents value and what accountability systems measure, and they highlighted the precarious political position that the accountability movement is in right now.

These are worth addressing in turn.

1. Lack of Clarity of Purpose

Tied into earlier concerns about the narrowing of curriculum and the privileging of certain facets of education over others is the confusion on the ground of the purpose of school accountability systems. On one level there is simply disagreement over what schools are supposed to be doing and thus what should "count" when holding schools accountable.

As one practitioner put it:

> "There's a lot more to forming a human being than just getting them to regurgitate an answer on the assessment. So, things around character, civic education, what does it mean to be human, what is human dignity, there's just a whole bunch of things that I feel like fell by the wayside, because they aren't measured, and it naturally fits in with the information age and programming things...so in order to know if you've successfully programmed something, you have to have a definition of what success is. That's how accountability becomes the definition of success."

But even once there is some agreement as to the types of data to be collected and the types of things to be measured, there is still confusion and conflict.

One policy advocate argued:

> "It strikes me that the current ecosystem is one of confusion, dissatisfaction, and frustration. So the confusion is that there seems to be a constant moving target to what it is these tests are, and what they're supposed to be doing. So I know in [STATE] we've had several different iterations of the state test. Everybody's frustrated about what that test is. What are our goals because the state board keeps moving the target. They keep changing the cut scores. They keep changing the tests. They keep delaying the high-stakes aspects of the test. So there's just confusion. There's dissatisfaction. Even the public school teachers and superintendents that I interact with, charter schools, they want more freedom. They want less regulation. They're frustrated by the top-down approach. So there's just tremendous dissatisfaction within the teacher community, the public school teacher community, the charter school teacher community, the private school. And then there's frustration because they want the rules to apply to everybody."

2. Layering

No raindrop is responsible for the flood. Schools develop their own assessments and accountability tools, as do districts, as do states. Each in isolation might make sense or serve a particular purpose, but adding them all together leads to a lot of time testing and a lot of competing mandates.

As one engaged outsider stated:

> "There's a district in [STATE]. Just last year, in one year, gave 79 tests that were mandated. Don't quote me on the 79, but it was something like this, 79 tests that were mandated, and the reason was that they just never went back and took away all the tests that were meaningless, right? And so we're just layering on all of these challenges."

A practitioner argued:

> "I think this whole idea of federal, state mandates to fix is really difficult for a school to absorb when a lot of these, they are all with good intent. Strategically at large, I think you'll see positive outcome, but then when you drill down again, each school, each community, each student needs something so different. So I think one of the misses is how we haven't figured out local autonomy for school accountability. It's really, really hard to do that. We haven't figured it out anywhere that I know of."

3. We're Not Measuring What Parents Care About

School accountability systems publicize math and reading test scores, graduation rates, and a set of indicators that policymakers hope that parents value. But do they? Many participants argued that parents are actually looking for different things (like school safety), and thus the measures we currently use are not as useful as proponents think.

It's hard to even identify the exact things that parents want to know, because parent desires differ. As one engaged outsider put it:

" I think in terms of what school information is most important, it varies. I think it depends on your family priorities. In some cases, safety is really important. In other cases, parents want their kids to go to a school that's going to inculcate them in a similar set of values to what's happening at home. Big emphasis on teacher effectiveness among some parents. The vast majority of families still aspire for their kids to have the opportunity to go to college."

One researcher agreed, contrasting current data collection to potentially different data collection:

" Is it what we, the experts, think they should care about? And they don't care about the same things. But I also think that's probably a mistake to lump them all in the same category. Parents want different stuff... Sometimes I want something else. Maybe my kids need something ... different kids need different things. But I think that's also part of the challenge. What's the view we should privilege? And if it's public policy, then maybe what's the goal for public good in a non-economist way. It's not the same goal as each individual parents."

All of this said, any centralized accountability system is going to chafe against parental demands, because parents are diverse and want different things out of schools. Any accountability requires some standardization and satisfying the hundreds or thousands of different visions of education that parents hold would require hundreds or thousands of individualized accountability systems. That simply might not be feasible.

4. Politics

Participants also highlighted several political problems that accountability policy has run into over past two decades, leaving it in a precarious position today.

One engaged outsider put it this way:

" [E]ducation isn't very savvy when it comes to dealing with politics. Whatever accountability system gets put forth, the next iteration, I think you have to think about the politics as much as the policy. If not, even more so."

A researcher said:

" It is very hard to get them to do precisely what we might want them to do through accountability incentives, and then I think another theme that I've heard a lot here that I think is right is that there are politics surrounding this—and understanding how different groups are going to react to different types of accountability is something that has to happen on the front end, and I think a lot of problems could have been anticipated and were not."

Another researcher said:

" I think the politics around testing matters. I think we've seen over the last few years a lot of perfect statistics, some of that has been parent opt-out movement, some of it has been teachers resisting, kids getting tested too long. I'm part of the group that is very drawn to the idea of a kind of inspectorate system that builds a qualitative component to supplement the kind of test-based approach. I think most of the argument for that is that people can see things the test can't see."

Another researcher put it this way:

" And, from a political standpoint, I think it's pretty clear that accountability doesn't really have a natural constituency. So, say school choice, one reason I think it's been relatively successful is, it provides direct benefits to people. Who really benefits from accountability? Not educators. Not parents, really. Potentially, they could, down the line. But, it's not very traceable. Taxpayers? They care so little about that. So, you're not gonna get a constituency that's gonna go to bat for accountability because no one really sees direct, tangible benefits."

THE FUTURE OF SCHOOL ACCOUNTABILITY

After taking several hours to discuss the past and present of school accountability, participants looked into the future. In their small groups, they had time to dream up their ideal state-level accountability system for a hypothetical Midwestern state. They then presented those preferences to the larger group and discussed the pros and cons of the various provisions they included.

Rather than reproduce those here, we'd like to cover several of the major themes that emerged in those conversations. Broadly speaking, they fall into one of three categories. The first is New Priorities, which look to the ways that accountability systems should evolve to respond to the lessons that we have learned and to better serve educators and families. The second is New Data, which is a list of all of the potential ways to measure these new priorities and better capture what parents want and what educators think is important. Finally, these forward-looking discussions opened up New Coversations About Old Questions. Even with some agreement about the problems of the past and present, participants struggled to agree on how to move forward. That discussion and debate, though, helped clarify where we might see disagreement and discussion over the next two decades of school accountability policy.

New Priorities

While our participants offered numerous new ideas for how school accountability systems could be operated, we'd like to highlight six. This is by no means an exhaustive list, but here are a provocative set of concepts to consider.

Priority #1: Restoring Trust

Foundations, business groups, and advocates need to rebuild trust with communities. One researcher brought the problem into stark relief:

> "I don't know, it just bothers me, we listen to these guys for a while and end up with schools that we hate, so they change their mind and they want to sell their stuff, you can't really do that either, right?"

One potential way to do this, as one engaged outsider proposed, was to focus on holding leaders accountable, not drive it down to rank-and-file teachers. As she put it:

> "What if you flipped it and created [an] accountability system that was just for leadership and measured that outcomes and if those leaders were able to meet those goals and the trust is that the students are as well?"

As one researcher succinctly put it:

> "People are becoming more and more distrustful of experts because they've overstepped."

So can those experts step back? Can they allow local communities to have more control over what accountability looks like for them? Can they put more trust in parents to weigh the various pros and cons of the school options that are in front of them?

If policymakers, advocates, and foundations don't trust parents and don't trust communities, they shouldn't be surprised when those parents and communities don't trust them back. Restoring trust and respecting people is step number one.

Priority #2: More Honesty About Tradeoffs

When a system chooses to privilege some metrics over others, those decisions have tradeoffs. Because schools have a finite amount of time and competing

goals placed on them by the various bodies that oversee them, they have to commit to certain things. That creates less time and space to commit to other things. To move forward, all parties in this system need to be more honest that each of those decisions has a cost. That cost may be worth it, but it is still there.

One practitioner put it this way:

> "We don't actually want schools to be accountable. We want them to be excellent. And that's a little bit of a duh, but I think, well how do you measure excellence? Well, you're accountable. You know, like you figure out these measures. But I think building on what [NAME] is saying, what you measure determines what you do. So I think, you got to start somewhere, and you know like, we threw this set of metrics at the wall, and we need to change it now."

Priority #3: Focus on Forming Human Beings

Participants believed that preparing students for the workforce is important. They believed that instilling basic knowledge of reading and math is important. But overwhelmingly they agreed that school is about so much more than that. One practitioner summed up this view well:

> "I think one of the things that's happened throughout the 2000s, and probably even before, is that we've begun to look at children as if they're programmable robots with the goal of, 'I'm gonna feed them this information. They're gonna spit out a result on an assessment, and then we're gonna call it a success.' And we've forgotten that actually as educators, we're in the business of forming human beings."

Part of solving this problem is changing the data that are used to measure how students are performing and how that data are used. Another part is treating children like human beings nested in families and communities and respecting their varied backgrounds, not just trying to wedge all of them into the same monolithic policies and institutions.

As another practitioner put it:

> "And so I think the way that we look at, the way that we look at developing goals for academic success I think needs to be looked at differently. I'm not sure what that would look like, but to me it's more of a looking at a wholistic ... at the whole child."

Priority #4: Building on the Strengths of America

When it comes to what accountability systems should focus on, and how they should sit on top of schools and communities, one practitioner offered a unique perspective:

> "Our strength in the United States is creativity and entrepreneurship, number of patents per person, things like that. So instead of focusing so much on our weakness, our deficit of testing in math, and NAEP scores, and PISA, let's focus on our strengths of entrepreneurship and creativity."

What would such a system look like? It would measure different things, certainly. It would also probably be structured different, have different goals, have different delivery mechanisms, and a host of other potential changes.

Priority #5: Continuous Improvement Over Carrots and Sticks

Teachers, school leaders, and parents want data. They want to know what is working and what isn't. But as soon as that data becomes part of a metric that measures school or teacher performance, it risks being corrupted.

As one engaged outsider put it:

> "Educators need information too. Parents need information to make decisions about whether they wanna stay in that school or not. But principals and teachers need information to have this continuous improvement cycle."

As one researcher put it:

> "Principals are gonna look at their suspension rates, they're going to look at discipline data. If one classroom is having much more trouble than another, what does that teacher need? There's a continuous improvement element of the data. But I feel like that's corrupted as soon as you tie it to accountability to the consequences."

During the large group discussion, one participant posited this:

> "So I don't remember, someone said this in our group yesterday, that the word accountability has a lot of baggage attached to it. I almost think that part of this is a reboot or a rebrand. We were kind of all toying with different terms yesterday like continuous improvement, transparency, parent engagement. I almost think to say that we're going to do something different, it almost needs to be a different word. It's come to mean just test-based accountability. How can that be changed at this point?"

Priority #6: Serving All Children and All Schools

Accountability policies have not affected all schools equally. This has caused political problems, as accountability is seen by suburban and rural families as something for other schools. It has also engendered pushback within urban schools that feel they are being unequally targeted when states in other communities are struggling as well.

One practitioner put it this way:

> "The accountability culture does not hit all Americans equally. The poorer you are, the more beholden to accountability structures you are. And while private school enrollment is dropping relative to total school enrollment across the country, I conjecture that both in terms of specifically opting out of testing when they remain in a public or charter school, and then opting into schools where they don't have the same accountability regimes, wealthy parents have a different relationship to accountability structures. And wealthy schools can have a different relationship to accountability."

A researcher echoed it:

> "But, I feel like, that's why...it should be kind of predictable, in hindsight, why people started complaining about No Child Left Behind was when the AYP went up enough that it started catching some of these suburban-ish districts with some of their subgroups. Then, they were like, 'Wait a minute, we don't like this anymore.'"

Accountability systems that are built to solve the problems of urban schools are not well suited to solve the problems of suburban and rural schools, and vice versa. Can uniform systems be designed that serve all types of schools in a given state equally well?

New Data

When developing new priorities, new measurements will be necessary as well. Participants generally coalesced around three design features for the types of metrics that schools and policymakers should collect in the future. First, those data should be measures that assess what matters. They should reflect what is really going on in schools and what families and educators care about. Second, they should be difficult to game. Campbell's Law tells us

that any measure that becomes used to reward or punish workers is going to get distorted, but there are some measures that are easier to manipulate than others.[17] Finally, they should be helpful and useful for educators and parents in both improving the work that educators do and improving the understanding that parents have of their child's education.

So what are some of these metrics? Over the course of the small group sessions, participants came up with a large number of them. It is helpful to think of them in subsets.

Student Characteristics

- **Family Structure**. Just like keeping track of student poverty and language learning status, states and districts should keep track of a child's family structure. Are they from a single or two-parent parent home? Do they have an extended family support network?

- **Portfolios of student activities**. School is about more than just academic preparation. A great deal of important socialization takes place in clubs, sports, and the school newspaper lab. How can we better understand the outside activities that students are participating in?

- **Chronic absences and chronic tardiness**. Do students show up to school, and do they show up on time? Understanding the attendance patterns of students could help understand their academic performance.

Student Academic Performance

- **Test scores beyond just reading and math**. Several participants called for a richer set of subject matter for examination, including science and social studies.

- **Longitudinal data**. Can schools collect data on college matriculation, remediation rates, college retention, college completion, employment, income, etc.?

- **Norm-referenced tests**. Several participants spoke of replacing one state test with the option of multiple norm-referenced tests that schools could choose from. One suggested making the PSAT universal and using that as the high school measurement.

School Quality

- **Waiting lists for schools**. Especially in areas with a great deal of school choice, schools with waiting lists are probably better schools. Schools with declining enrollment are probably not so good. This allows parents to the be the eyes and ears of those trying to measure school performance.

- **Course offerings**. Do schools offer AP classes? Career and technical options? Parents could use this information to decide if the school is the right fit for their child.

- **Micro-credentials and industry certifications**. Are students leaving school with credentials recognized in the workplace that can help them get jobs? Particularly for students who aren't college-bound, are they getting the recognition they need for the technical training they have received?

School Climate

- **Student and teacher retention**. Are students and teachers voting with their feet? Schools that see a lot of turnover in their students and teachers probably have issues that need to be addressed.

- **University of Chicago Five Essentials school climate surveys**. This particular evaluation tool (which measures "effective leaders, collaborative teachers, involved families, supportive environment, and ambitious instruction") came up repeatedly as a means to evaluate school climate.[18]

- **Parent satisfaction surveys**. Many businesses use consumer surveys like the Net Promoter Index.[19] Are there tools that could be used to measure parent satisfaction that are user-friendly and useful for educators looking to improve their practice?

- **Climate and engagement surveys**. More extensive surveys of students, parents, and teachers were floated as tools to better understand what is going on in schools. One participant emphasized that these tools would need to be anonymous to guard against gaming and/or reprisals for critical feedback.

- **360 evaluations**. Particularly when it comes to evaluating teachers and administrators, this common human resource tool came up as an option. It encourages self-evaluation and the gathering of feedback from multiple peers as well as managers to get a holistic understanding of an individual's performance.

- **Arrests in school**. An admittedly blunt metric that could be gamed by discouraging 911 calls, arrests and/or 911 calls from schools were floated as a measure of school safety.

- **Composition of staff beyond years of experience and credentials**. For parents hoping to choose a school, participants wanted to know more about who was teaching in the school. What is the racial composition? What are their strengths and weaknesses? What is their background? Are they from the same community? Do they have life experiences that could have a positive influence on students?

Policy Metrics

- **Stricter accountability measures at fewer checkpoints**. Across the country, students are assessed in reading and math in grades 3–8 and then again once in high school. Throughout the small group sessions different participants offered different potential arrangements with fewer tests (say at just fourth, eighth, and tenth grade). Some wished to pair those fewer exams with more strict checkpoints for schools and students, such that students would not progress through each checkpoint without demonstrating that they were where they need to be.

- **Return on investment**. Several participants wanted more care toward and measurement of return on investment. Are new initiatives getting bang for their buck? Perhaps they were successful but at great cost? Perhaps they had small gains but also little cost? Understanding both sides of the ledger can help policymakers and community members in evaluating how to judge program success and how to best allocate resources.

It should be noted that these data elements were not all deemed equivalent by participants. Some thought certain elements were far more valuable than others, but all are listed here to try and spark conversation and thinking about potential new data sources in the future.

A Deeper Discussion about Data

There also were two areas where affinity groups dug in deeper that are worth expounding upon. The first is around how states and districts might go about identifying and working with low-performing schools. In several instances, participants argued for a more qualitative approach to tackling the problem.

One researcher's thoughts are worth quoting at length:

> "Yeah, and a lot of this is about kind of gaming and cheating, and to me, the most puzzling part about the American approach to accountability is that we try to do it cheap. We do not, for what I think drives just about every decision schools make, which is our accountability systems, we try to do it really inexpensively and part of that decision is that we have opted for an almost entirely quantitative evaluation approach.
>
> I am a quantitative researcher. I believe in quantitative methods. Having said that, when it comes to evaluation, other countries do a lot more qualitative work to accompany the quantitative work. They send inspectors, they have eyes on classrooms. They look at lesson plans, and at scope and sequences. They have some sense of what's going on in classrooms.
>
> We have focused on metrics like attendance rates, graduation rates, test scores. Things that are really easy to measure, and things that are really easy to measure like that are often much easier to game than trying to sort of wiggle your way through having some expert eyes who are taking a look at what's going on in the school.
>
> What I often hear as an explanation is like, 'Yeah, that other stuff is really expensive.' I don't know the numbers on it. My suspicion is that in the sort of grand scheme of things, it would be a drop in the bucket to invest more in accountability relative to the impact it's having on what schools do."

The second area of deeper discussion is around measurement of socio-emotional learning (SEL). SEL has become a buzzword in education, and there have been calls to try and work SEL measures into school accountability systems. Participants disagreed as to whether or not that was a good idea.

One engaged outsider argued for including socio-emotional learning like this:

> "You still need to be able to assess the foundational skills of reading, writing, mathematics and others. You also need to be able to identify and evaluate a set of other social and emotional learning skills. While many of us talk about it, the assessments to really measure our progress needs a lot of work, and now states are being asked in ESSA to do that."

Another engaged outsider raised challenges:

> "Which set of the range of social and emotional development needs? What are the priorities? What are the set of skills that are most connected to the success that we want kids to have in life. I'm agreeing with everybody. It's something we want to do, but there are challenges in doing it well."

Yet another engaged outsider was blunt in her assessment:

> "No. Do not create accountability for SEL, please."

New Conversations About Old Questions

Participants were not able to resolve every difference of opinion, nor did we expect them to. In disagreeing and debating, they did offer several questions around which future conversations about school accountability might swirl. Some of them are new versions of old questions made salient by 20 years of accountability experience or by the new mandates and expectations of ESSA. Others are new questions prompted by evolving values and new methods of teaching that will have to find ways to integrate themselves into accountability systems. We'd like to highlight five questions that participants raised that we think will shape future discussions of school accountability.

Question #1: Is the point of accountability to establish a minimum standard or to drive schools to be better?

There was robust debate about the ultimate purpose of school accountability systems. This is not a new debate, but one that even after 20 years is still unresolved.

One researcher sketched the outline of the debate this way:

> "I mean, I'm inclined more that way than trying to argue that we're gonna use this system to somehow cause schools to be better. Because I don't know how schools are going to be made better. Are they going to be made better in a lot of different ways? So if we want to define accountability, I would kind of incline towards, let's not just light taxpayer money on fire. Right? So you're meeting some minimum standard, we'll leave you alone to do all the stuff you think you need to do."

One group of practitioners spent some time discussing "guard rails." This is one way of looking at the goal of the system: to provide some basic levels of achievement or basic goals that schools need to meet in order to stay in operation or receive public funds. But where those guard rails are set matters, are they supposed to be minimum standards, or are they rigorous ones? There was not agreement across participant groups.

A policy advocate raised this concern:

> "It always seems like we're creating...the floor and saying 'this is the bare minimum you have to do, but you can do so much more.' We end up creating, more times than not, sadly, I think, a ceiling. That's all that people go to. 'I have to do. I have to report this stuff. We have to do these measures, so that's what we're gonna do, but we're not gonna do any more than that, but we have to do those.'"

Can guard rails really be a bare minimum? Can they drive rigor? Opinions diverged.

Question #2: To grade A–F or not to grade A–F?

Rolling up school performance data into a single metric, in some states A–F, in others a kind of cumulative performance rating, was a topic of discussion in several affinity groups. Opinions were mixed on this practice.

One researcher summarized the debate like this:

> "How do you make something that is simple enough to be understood, like an A through F rating system, but also incorporate a number of different factors that are complex enough to capture all of the things we want schools to do? Everything from math and reading to also discipline data or enrollment data or attendance data or all these other sort of facets of that system. So how do you make something that is usable and understandable, but also nuanced? And I feel like there's just, we're all just sitting here against a wall in that sense."

For the idea of A–F grading, one engaged outsider argued:

> "Again, [STATE] isn't perfect by any means, but just the policy landscape has really enabled the utilization of A through F to actually empower parents to make better decisions about where their kids go. I think that's the goal."

Against the idea of A–F, one researcher said:

> "I sort of feel like the single rating of either A through F or on a number is sort of the worst impulses of accountability. Because not only are you saying what matters by its inclusion in that, but how much it matters, by how it's weighted. So man, that takes a lot of faith in yourself that you can specify how much you should care about academics relative to attendance, relative to these other things."

Clearly, there are competing values to weigh here. Giving families and community members clear and straightforward information about how schools are performing is something that many participants wanted to promote. But others had misgivings about the data points that made up the constituent parts of those ratings, how those data points would be weighed against one another, and what might be lost when they are all rolled up into one measure.

Ultimately, it comes down to a question of comfort with complexity. Are we comfortable with complexity? Do we believe parents and community members are comfortable with complexity? How can we better understand how parents and community members consume and act upon information about school performance?

Question #3: Can we have strong accountability without enforcing uniformity?

Much time was spent discussing and debating the effects of school accountability systems on schools. One concern that surfaced multiple times was the risk of homogenizing education that is inherent in trying to apply a single set of standards to schools. Some folks were more comfortable with the tradeoff than others.

One policy advocate put it this way:

> " I think personally from my lens, it's the crux of the whole conversation we're having. Can we have accountability, strong accountability, without it having to be uniform? "

Part of this is a problem of definition, as another policy advocate put it:

> " What does uniform even mean? That's in our constitution, our state constitution, uniform. When you say 'uniform' do they mean equal? I wish we could just say what we mean. "

Another policy advocate summarized the debate like this:

> " Yeah, we keep coming back, at least my read and maybe it's just my lens right now, how do we, if we all value a pluralistic approach to education, with that in mind, how do we define success? And then with that in mind, how do we measure that from some sense of accountability? And I don't think we can do it in one uniform way, and that's what our government structures try to force us to do is to have one uniform way to measure success when success is so diverse, and we want to have diversity of options out there... so every child can find that right place that's the right fit for them so they can be successful. "

During the large group discussion, one participant said:

> " One of the things that came up in our conversation multiple times is the trying to promote plurality in educational options, but that coupled with maybe, the challenge of a uniform accountability system, and is that even possible to come up with a single accountability system that would address the diversity of options that we're trying to create and implement across the country. "

This tension is not going away and is something at the top of the mind of many of the people who participated in the small group sessions. It is not something that advocates or policymakers will be able to cast aside. We must wrestle with this moving forward.

Question #4: Do we trust parents more than the school system?

In broad terms, almost all of the participants in the focus groups believed that we should trust parents to make decisions about the best education options for their child. However, when discussions got down to the nitty-gritty details of what should be allowed and what shouldn't, fissures started to emerge.

One policy advocate raised this question directly:

> "There's a huge assumption that we're ignoring, and I deal with that on a day-to-day basis at my job: Do we completely trust parents to make the right decisions for their kids' education?"

Another policy advocate made the case for parental supremacy in this way:

> "These are the people that are raising these kids, they're responsible for these kids, from the minute a kid is born you become their number one advocate, you're making all these decisions. And then why is it once they enter public or private school … all of a sudden you aren't the expert any more, and they know better? There's this crazy dichotomy. We hear from the schools, 'We wanna engage parents and we want parents to be behind us', and then the minute a parent says 'Well, what about this?', all of a sudden it's the crazy emotional mom or dad or whatever it is, and then the door shuts."

But a practitioner raised a point about the numerous stakeholder groups in education that schools must answer to. Privileging parents could set schools in opposition to the wants and desires of these groups:

> "When you think about student, teacher, parent, educator interests, I think there is a heightened concern about accountability for learning. I think when you start to get to the business community, to higher ed, to policymakers, I think learning is important, but it also becomes about things like the public trust and about ROI, and about … 'are we getting what we are paying for?'"

A policy advocate put it directly:

> "Sometimes there are some pretty bad schools out there that need to close."

For folks who agree with this sentiment, the question "by whom?" looms large. What government entity should do this? Should it be local districts, the state, independent entities, or someone else? Some groups also asked if there are sanctions for schools that fall short of closure that can help spur schools to improve. Perhaps a nudge (or a push) shy of closure can help schools improve.

This question is not going away. Even within the coalition of school choice supporters, there are differing levels of support for parental supremacy in decision-making. There is no clear resolution to this in sight.

Question #5: What should be the unit of analysis? The district? The school? The teacher? The learner?

Accountability systems have historically focused on schools. Children increasingly are getting their education outside of school. Schools also are reconfiguring to offer more student-centered options around the classes that students take and the rate they are able to progress through them. Policies like Education Savings Accounts and Course Access programs allow students to piece together their education from several different providers. All of these chip away at the foundation for school accountability. What does accountability look like in this new world of educational provision?

One researcher summed it up like this:

> "But it's no longer clear what the school is. And that matters a lot for accountability…it used to be very simple. The school building was the same as the unit of accountability was the same as an organizational entity. No longer, that is no longer the case when you look across districts and states. This is no obvious definition unit of analysis. And it really matters for accountability."

Another researcher responded:

> "How do you evaluate those responses? What was the effective part of that combination of treatment mechanisms?"

One engaged outsider made the case that the learner should be the unit of analysis:

> "I think the shift is now to the actual individual learner. I think focusing on the school as a unit of change is problematic...I think if learners are gonna be successful navigating today's world, they have to develop their own agency to be able to do that, and every learner is different, and we don't have a standardized learner, and we try to push every learner into a standardized system, so the push towards more personalization and customization and building learners' journeys to life is gonna require not just going to school. It's gonna require a variety of different learning experiences. It's gonna require some, more or less, emotional support, depending on the personal context of that kid. It's challenging, and I don't know exactly how we're gonna do it, but when you have education savings account and tax credits, where kids are going to different places, we're not living in a society any more that is standardized in that way."

Trying to evaluate programs at the learner level is much more challenging than at the school level, but it is not impossible. This would require a retooling of school accountability systems, and serious work trying to parse out the effect of different educational providers on a child's learning.

IMPLICATIONS

After reviewing more than 40 hours of video of the large and small group discussions and combing through the transcripts, which total 790 pages and some 300,000 words, it feels like we just finished trying to drink from a fire hose.

That said, we would like to offer some editorial thoughts about what these conversations have to say about the future of educational accountability. These are simply our (McShane and DiPerna) opinions after observing the affinity groups, and they obviously are tempered by our own biases and experiences. To be perfectly clear, they in no way should be seen as representative of the thoughts or opinions of those who participated in the focus group sessions. They are a reaction to them.

1. A diversifying educational ecosystem is an existential challenge to traditional notions of school accountability. Maybe this is for the best.

Our educational ecosystem is diversifying across several dimensions. More and more traditional schooling options are becoming available for students. In many major metropolitan areas, students have multiple schools to choose from in the traditional public-school sector, from open-enrollment schools to magnet schools to specialized schools. By and large, most metropolitan areas also offer at least some charter schools. A small but growing number of metropolitan areas offer access to private schools with public support. It is more of a mixed bag in suburban and rural areas, but increasingly those schools are finding opportunities in the second dimension of diversification, the "unbundling" of education.

Unbundled education sees families combining the offerings of multiple education providers to educate their child. They might use an online math program paired with in-person tutoring. They might use language-learning software or therapy services offered by local non-profits funded by an Education Savings Account. There is no one "school" that the child attends. Or the school is really just a clearinghouse for the providers. Outside entities are increasingly doing the work of educating children in these environments.

Both of these kinds of diversification threaten school accountability as we know it. Without a single entity to "own" children for accountability purposes, A–F letter grades or performance matrices are harder and harder to construct with validity and legitimacy. Different providers might have different levels of effectiveness, so how does that roll up into a holistic measurement of how well a student is performing, and who do we hold "accountable" if that performance is not up to snuff? Choosing from a variety of options also lessens the need for heavy-handed centralized accountability, as families are able to play a more active role in quality control by voting with their feet. A diverse set of school choices also challenges us to think about what metrics might measure schools with different philosophies and how we can compare schools with different pedagogy and goals to each other.

But families are not the only stakeholders in the education system. Taxpayers are footing the bill in most cases. There is strong public desire for those who receive public money to be held to account, and this is not going away, nor should it.

The question is, are centralized accountability systems using a relatively narrow and noisy set of metrics the best way to hold schools and educators to account for their actions? We would give a qualified "no." That is to say, we think that having central systems for collecting and disseminating performance data on schools is important, and the set of metrics that are collected and disseminated should be diversified and expanded, but ultimately, those metrics should only be used in rare cases to override the desires of parents. Primarily, they should be used to inform decisions, both the decisions of parents and the decisions of taxpayers, they should supplement human judgment, not attempt to replace it.

Our recommendation is that accountability adhere to a general principle of restraint by prioritizing the collection and dissemination of information that is valid, reliable, transparent, relevant, timely, and maintains clarity for the intended stakeholder or audience. If a new potential data point cannot satisfy those criteria, if it is just "nice to know," it should be excluded.

2. Accountability advocates have a long way to go earning parents', teachers' and the public's trust back, if they ever will.

One of the things that surprised us the most about participants in our affinity groups was their overall pessimistic view of accountability. We expected some participants to be negative about current state accountability frameworks and systems, but not the overwhelming majority. Asking about the mistakes and flaws of accountability systems in the last 20 years opened up the flood gates. Numerous participants clearly had changes of heart as these policies were rolled out. They were honest about their overconfidence in accountability's ability to improve schools. They were much more modest in their ambitions today.

This mirrors views on accountability from society as a whole. We have observed similar shifts in public opinion in the direction of declining support first for No Child Left Behind, and then again for the Common Core State Standards (and its accountability implications). When it comes to testing, on the one hand there is ample survey evidence showing that the public feels there is too much testing. [20] But on the other hand there are also public opinion data showing that majorities of the general population believe standardized assessments should be used for accountability purposes. Interpreting public opinion is complicated and downright messy when it comes to the topic of accountability in K–12 education. The mixed signals we have so far further reinforce the call for humility and caution when thinking about designing accountability frameworks and attaching carrots or sticks.

For those who are still accountability hawks, priority number one has to be winning back the trust of parents who don't value what most

accountability systems measure, teachers who feel maligned by test-based accountability systems that weren't ready for prime time when they became a part of accountability frameworks, and a public who has lost faith in the value of standardized testing.

It is not immediately clear to us how they can do this.

3. There is a lot of agreement on the problems of school accountability. There is less agreement on how to fix it.

Relatedly, our participants agreed a great deal on the shortcomings and problems of the last 20 years of accountability. The same problems came up in group after group: test obsession; reading and math obsession; a shunting away of other important goals of schooling; as well as alienating teachers, alienating parents, and gameable metrics. There were varying levels of agreement, but by and large, folks agreed that these consequences were unfortunate and largely driven by accountability policies.

When the task moved on to trying to determine a path forward, there was far less agreement. That was our clear observation when we had affinity groups create a hypothetical accountability system for a midwestern state with a diverse set of educational options. What metrics are important? Are we shooting for a baseline of performance or trying to promote excellence and continuous improvement? Does this process need to be more qualitative and include more human judgment? Is transparency enough? Is school choice in and of itself the best form of school accountability?

There were real divisions here, and again, not a clear path out. Insofar as states are looking to create a uniform system, it will be hard to placate these competing concerns. Perhaps moving to a system where accountability is seen more as informative— that is, providing clear and transparent measures

of school performance on a variety of metrics that parents and taxpayers can rank and value based on their own beliefs—is the best way to thread the needle. Leave it up to parents to weigh if higher math scores are worth lower ratings of school safety. Leave it up to taxpayers to decide who to elect to the school board or whether or not to support a millage levy based on this available data. Don't try to make decisions for them.

4. There is a huge degree of path dependency (or Stockholm Syndrome) when we talk about school accountability.

Participants were given a blank slate to create a new accountability system for their fictional Midwestern state. In a presentation before the conversation started, we emphasized that the sky was the limit and that they should feel free to think about new and different metrics, structures, tools, and processes.

By and large, the accountability systems that groups created were roughly similar to the accountability systems that we have in states today. Sure, there was a tweak here and a tweak there, some different metrics here and some different processes there, but mostly it was some combination of student test scores, later life outcomes, and school climate indicators.

We can only conclude that contemporary conversations about accountability in K–12 education appear to suffer from some serious path dependence. Current notions of accountability, like testing kids in grades 3–8 and measuring climate via surveys and tracking graduation rates have become part of the grammar of schooling, and it is hard to think about accountability systems that are not based around these features.

To be clear, throughout the small group sessions, more radical ideas were proposed, and we did our best to highlight some of those ideas in the body of this paper. When folks came together, though, the

consensus moved back toward familiar concepts, goals, processes, and outcomes.

5. The search for non-gameable high-stakes metrics continues. It is unclear if such a destination is reachable.

Perhaps one reason why new proposals were so similar to old proposals is because it is difficult to find metrics that are hard to game. Our convening of affinity groups occurred around the same time that Ballou High School in Washington, D.C., was in the news for graduating students with extreme numbers of course absences and pressuring teachers to pass students despite poor performance. [21] The efforts that schools have taken to distort accountability metrics was clearly fresh on participants' minds. The problem is finding metrics that were harder to game.

Test scores can be gamed. Graduation rates can be gamed. College matriculation rates can be gamed. Surveys can be gamed. Counts of arrests or 911 calls from schools can be gamed. Waiting lists can be gamed. The list goes on. Sure, some are easier to game than others, and the level of fraud and coordination that it would take to game some is much higher than to game others, but given high penalties attached to these metrics, there is strong motivation to work around them.

Qualitative measures appear much harder (though not impossible) to be gamed. But, qualititative measures like school inspections or student portfolios of work are much more expensive to collect and are inherently much more subjective. If we cannot agree on what the goals of schools should be, do we want to outsource that to inspectors, accreditors, or evaluators? Perhaps they can better respect a pluralistic vision of what our education system could look like. But such a system also runs the risk of far more opacity that cannot be systematically examined or evaluated itself. This is not to say that such a system might not be better, or that it should not be in consideration as an alternative way of holding schools accountable. It is just to say that we should be clear-eyed about the tradeoffs.

Given this, it is hard to not take a nihilistic approach toward most accountability systems. Or, at least, toward accountability systems that privilege a few metrics and make serious, consequential decisions based on them. A workaround might be using observational metrics to inform consumers or to trigger more holistic evaluations performed by human beings with the ability to balance quantitative, qualitative, and subjective information for more substantial judgments.

CONCLUSION

This project is not the last word on accountability in K–12 education. In fact, we hosted these convenings in the hopes of sparking a new round of debate and discussion around how we talk about accountability. Who or what is being held accountable? Who or what should be held accountable? By whom? To what end? How should we measure performance, and what we do with that information? How you answer those questions has serious implications for what features should be included in your preferred accountability system. As these debates continue, it will be important to keep in mind the questions that our participants wrestled with throughout their time in their small group sessions.

We would be remiss, though, if we didn't conclude with calls for humility and restraint. Schools are complex organizations, and our school system is a complex network of governmental and non-governmental actors. The tens of millions of people who either work for or send their children to school have myriad opinions about what they want from those schools, what schools should look like, and what values schools should promote. Trying to find one system, or even 50 systems, that can satisfy those diverse desires is a tall order. Trying to find the five, 10, or even 20 metrics that can capture

what we want from schools is a tall order, as well. Removing human judgment and trying to rely on objective measures may very well be a fool's errand, as the determination of which metrics to include or not include is a subjective process in itself.

Perhaps recognizing that accountability has limited potential to drive improvement in schools is the right course. It might even be true that the gains that we would expect to see from accountability systems have already been captured over the past 20 years and we need to move on to new policy areas if we want to build on them. It will be some time, if ever, before researchers will be able to answer that question.

But our school system would be well served with a better understanding of the following questions, something that researchers, policy advocates, practitioners, and engaged outsiders could come together to help better understand:

1. What are the metrics that parents care most about? Do these metrics vary by location, demographics, age of children, or other factors?

2. What metrics do taxpayers care most about? Do these metrics vary by location, demographics, age of children, or other factors?

3. How do the various metrics that have been proposed correlate with each other? That is, do we see high levels of connection between schools with low reports of violence and higher student performance? Are there metrics that don't correlate? How do we think about what those tell us about what is going on inside a school?

4. Are parents successfully closing down schools that perform poorly on traditional measures of accountability in areas of robust school choice (that is, can parents obviate the need for centralized accountability systems)?

5. How do regulations shape who participates in school choice programs, either on the side of schools choosing to participate (or not) or families choosing to participate (or not)?

6. How do school-centered networks of teachers, administrators, and parents amplify or undermine the intent of accountability implementation coming from the school board or state government? How do social/professional networks affect accountability systems by allocating and reallocating information, power, or resources?

7. Are taxpayers willing to invest more money in collecting more measures or potentially hiring many more people to more subjectively evaluate school performance?

Until then, we should continue to debate, discuss, and experiment in this field. More learning is necessary, no matter how you measure it.

Beth Blaufuss

Independent Consultant, Partnership for Inner City Education Former President, Archbishop Carroll High School, Washington, D.C.

As Mike McShane and Paul DiPerna note, our focus groups faulted testing-driven accountability for narrowing educators' focus on too few results. It did not need to, though. That's where two topics for further exploration emerge: not metrics themselves, but our beliefs and motivation surrounding metrics.

Embedded in high-stakes testing is a belief that simply knowing accurately how we are doing can propel improvement. Did that work? Yes: progress accelerated for all kids under NCLB, and it improved faster for kids of color.[i]

Why didn't we throw a parade? I suspect two reasons: we've still got a long way to go, particularly to close achievement gaps we see more clearly; and the way we used data demoralized and stressed us out.

In addition to being an educator, I've also been a member of group that has taught me more than anyone about the impact of data: Weight Watchers.

If you've ever stood on a scale after exercising, pretending carrots are as tasty as brownies, and drinking enough water to consider moving your office nearer the bathroom, only to find the scale number barely budges—you understand. Weight Watchers knows their entire business model is based on helping people see steady progress and motivating them when it stalls. They can't change the number on the scale; they can change how you think about it. Perhaps we should have had a few of their psychologists with us at the EdChoice event.

As McShane and DiPerna report, we suggested new metrics, but we didn't really discuss revolutionizing accountability. I wonder if that's because we didn't dig into the beliefs implicit in current structures. In addition to believing data motivates improvement, we also implicitly believe that data and cause are easy to link. If I didn't lose weight one week, I must have done something wrong. If a teacher's kids don't improve, it must be the teacher's fault. We should both be held accountable.

The American dream embeds in us the notion that individuals are in control of their destinies. Failure, then, can produce shame. Shame motivates some. It makes others of us stress eat, deflect blame, and lower the bar for success.

What would shame-resistant accountability look like? The schools I work for—the Partnership for Inner-City Education—have beliefs that point in that direction. Data are neutral. Context matters. Interpreting facts deserves thoughtful deliberation. Good data is honest data; high scores achieved by compromising values isn't good data. Good action plans based on data interpretations involve calculated risks; risks are hard and should be valorized. Stretch goals are insufficient. Our evaluations and compensation reflect these beliefs.

Finally, we may believe testing data can tell us something it only partially reveals: that our kids will be OK. Educational improvement is a necessary but not sufficient condition for securing the American dream for our kids; accountability is only one part of that improvement. Let's be as smart about holistically seeking that change as we hope our kids will become.

---

[i] Lanae Erickson and Stephenie Jackson (2015, February 6), Did No Child Left Behind Work?, retrieved from Third Way website: https://www.thirdway.org/memo/did-no-child-left-behind-work

Tillie Elvrum
Parent Advocate and
President, National Coalition
for Public School Options

As a parent advocate, I believe a reevaluation of the current school accountability system is long overdue. There is a fundamental difference between why parents choose schools and how education technocrats evaluate school quality. When we evaluate and hold schools accountable primarily based on standardized test scores we lose sight of the reasons parents employ school choice including school safety, curriculum and academic programs, and student behavior and attitudes toward learning. Parents understand the need for testing, but it's often not timely, meaningful, or actionable to students and their teachers. Our reliance on standardized test scores to determine school quality has led to schools gaming the system, parental frustration, and policies that are stifling innovation and undermining informed parental choice.

Parents understand an A-plus school won't always be the best fit for every student. Authorizers should empower parents with choices and evaluate schools using multiple measures of performance. Ultimately, we should trust parents with keeping schools accountable.

Dale Chu
Independent Consultant
and Senior Visiting Fellow,
Thomas B. Fordham Institute

Within the reform community, a tension that has been brewing under the surface for years is now playing out in full view. It pits those who believe that parent empowerment (i.e., the power to vote with one's feet) equals accountability versus those who subscribe to the "grand bargain" of accountability for results.

I attended a workshop this year that endeavored to bridge the divide. Hosted by EdChoice, the event brought together a diverse group of stakeholders, including researchers, policymakers, practitioners, and philanthropy. There were folks in attendance who were fans of the Bush-Obama era, as well as those who felt the reforms implemented under their tenure were a detriment to students and the system writ large.

Participants wrestled with a number of key questions, such as: What is the purpose of an accountability system, and who should be held responsible for owning and operationalizing it? Though most folks felt that recent history has been too top-down in this regard, finding the sweet spot between local autonomy and technocratic overreach proved elusive—at least for the limited amount of time we had together as a group.

Complicating matters further is that the A-word has become a dirty one now synonymous with standardized testing. Certainly accountability means so much more, but it's reasonable to argue that the well has been poisoned at this point. "Transparency" and "responsibility" came up as potential alternatives, though how to measure either would also require extended discussion.

There's also a debate about what test scores tell us about school performance. To be candid, my thinking on the subject has evolved over the years. I still believe in the relationship between short- and long-term outcomes, but I'm more open-minded now to schools that deliver positive life outcomes without the stellar test scores to match. I'm also more appreciative of the intangible elements that make school and learning special, but may not lend themselves easily to hard measurement.

That said, I see no reason to ditch test-based accountability anytime soon. I still firmly believe that the problem with standardized testing is less about the complaints (e.g., they're biased against some students or too easily gamed) and more about the lack of outrage at the number of schools—especially those serving low-income

communities—that are not teaching the basic reading and math skills necessary to pass them.

Still, there seems to be a lot more questions than answers on the issue. It's all enough to make an honest policy wonk nervous that even if we were able to build the ideal accountability system, there's a ceiling to what it can accomplish on its own. But this doesn't mean abandoning accountability. Instead, we need an evolution that reflects upon what we've learned over the last decade, prompts further discussion and debate, and approaches the development of new systems with empathy and humility. [ii]

STATUS PANIC
Robert Pondiscio

Senior Fellow and Vice President for External Affairs, the Thomas B. Fordham Institute and Senior Advisor to Democracy Prep Public Schools

From time to time, the ground shifts beneath the feet of influential, credentialed, and elite professionals with well-defined sets of skills and interests, who collectively labor under a set of assumptions about the way the world works and their role in it. It's an odd and uncomfortable thing to witness when it happens. Call it "status panic." When a group of people who have grown accustomed to a level of deference and outsize importance in their sphere of influence feel it slipping from their grasp, some lose the ability to adapt and course-correct. Their ideas about how things must be may be unpopular, ineffective, or both, but they can't or won't accept it. They start to behave unpredictably as people gripped by panic are wont to do.

The most vivid and visceral example of status panic has been driven by our national politics since the 2016 presidential campaign. Thought leaders in the media initially responded to the rise of then-candidate Donald Trump with amusement and condescension. When his campaign didn't whither in the glare of their collective derision, dismissal hardened into revulsion, then full-throated and increasingly hysterical alarm, a phenomenon best illustrated by political pundit Andrew Sullivan's cover story for New York Magazine, who wrote that Trump's election would be "an extinction-level event" for liberal democracy and constitutional order in America. [iii]

If Trump's election was an extinction-level event for anyone, of course, it's Andrew Sullivan, political professionals, think-tankers, and other "opinion leaders." If you throw the full weight of your condemnation against a target to no effect, when those in power have achieved their status not just without you but despite you and owe you nothing—your options are limited to a sober re-evaluation of your theories of change and your professional impact; or shaking your fist and shrieking ever more shrilly into the indifferent wind. Many, alas, mourning for their lost status, seem incapable of doing anything but the latter.

To be clear – and grateful – the world of education policy, even at its most self-absorbed and contentious is genteel compared to our national politics and media. But creeping status panic was very much in evidence at the EdChoice convening in Ft. Myers, Florida. I was invited as an observer, not a participant. I left feeling like I'd just spent three days listening to several of those present—mostly those whose view of school accountability means testing (while agreeing, of course, that ill-defined "multiple measures" were important) trying to persuade themselves that their ideas still mattered, and that their status remained intact and undiminished.

[ii]An earlier version of this commentary was published by the Thomas B. Fordham Institute. Dale Chu (2018, June 20), Nervous About the "A" Word [Blog post], retrieved from https://edexcellence.net/articles/nervous-about-the-a-word

[iii]Andrew Sullivan (2016, May 1), Democracies end when they are too democratic, New York Magazine, retrieved from https://nymag.com/daily/intelligencer/2016/04/america-tyranny-donald-trump.html

This is an uphill argument. As I have argued elsewhere, education reform's policy agenda over the last several decades has transformed American education in ways that have proven unpopular and ineffective. If the standards and accountability movement was going to usher in the promised rising educational tide that lifted all boats, the benefits would outweigh the disruption and be worth continuing the fight. But if it hasn't happened by now, it's probably not going to happen at all.

This is not to say that there haven't been important victories. Accountability hawks deserve no small measure of praise, as Mike McShane and Paul DiPerna observe in their summary of the convening, for "increased focus and attention on educational equity and the yawning gaps in achievement between different student subgroups, greater transparency, improved data literacy," and other benefits. These have been clear wins for students and American education at large. But those victories have come at a cost: curriculum narrowing, excessive test prep, an ever-increasing number of hours and days of instructional time thrown into the voracious maw of mandated tests, to say nothing of deleterious effects on the culture of schooling at large. A significant number of children surely feel that their school's primary purpose is to prepare them to sit for annual tests in reading and math. And their parents have clearly had it. To put it mildly, it has been challenging for ed reform's testing technocrats to accept the discontent is real and enduring, not the mere whining of parents upset to learn that their children's schools are poor performing.

There is an unexamined perversity at the heart of test-based accountability and always has been. The logic model assumes that districts, schools, and teachers are well versed in professional practice and competent to deliver it; they merely need to be incentivized—held accountable—for doing so. The historic trend lines of NAEP make it clear that American schools were stuck in the doldrums before "A Nation at Risk." They are largely still becalmed, three decades into the standards and accountability movement. Likewise, it won't do for reform critics to suggest that educators are capable of delivering quality instruction but are hamstrung by accountability policies.

To be fair, the cognitive dissonance on display at the EdChoice convening—the desire to do something but having within one's reach only tools insufficient to the task—may be an unresolvable conflict. As long as schools are run primarily on public dollars, there will be demands for public accountability. This has become the de facto battle cry of technocrats in status panic mode: we can't merely trust the teachers and send more money. Nor will it do to let parents vote with their feet. We must ensure that every choice is a good one! For now, the forlorn and unspoken hope of accountability hawks and testing technocrats still appears to be that somehow, someday, parents will value test scores as much as they do.

Technocracy dies hard. Whether it's borne of a distrust of schools and parents, a stubborn belief in one's own righteousness, or a simple inability to look outside and take note of the change in weather, those in the grips of ed reform status panic are at risk of marginalizing themselves through a stubborn insistence that "Well, testing isn't perfect. But it's better than nothing." There are many other stakeholders in public education. And many other ideas about accountability.

APPENDIX 1
Qualitative Research Profile

PROJECT TITLE: Working Toward a New K–12 Education Accountability Ecosystem

DEVELOPERS: EdChoice, Hanover Research

FACILITATORS: Hanover Research

FUNDER: The Walton Family Foundation

POPULATION: Professionals with interests or who work on K–12 education issues

LOCATION: Ft. Myers, FL

INSTRUMENTATION: Focus Group protocol

DATA COLLECTION STRATEGIES: Moderated group conversation

LANGUAGE: English

SESSION DATES: April 24–25, 2018 (Convening #1)
April 26–27, 2018 (Convening #2)

DATA COLLECTION LOCATION: Hanover Research, 4401 Wilson Boulevard, Arlington, VA

DATA COLLECTION METHOD: Live, in-person meetings

SESSION LENGTH: Three Focus Group sessions were 75 minutes
Three Convening sessions varied, 45–75 minutes

PARTICIPANT ELIGIBILITY: Invitations based on professional interests or work on K–12 education issues

TOTAL NUMBER OF FOCUS GROUPS: 8

TOTAL NUMBER OF PARTICIPANTS: 47

STIMULI: N/A

TRAINING OF PROFESSIONAL FACILITATORS: Yes

SUBJECT COMPENSATION: Each participant was offered $1,000 honorarium. All travel, meals, and hotel were paid on participants' behalf.

AUDIO OR VIDEO RECORDING: Yes (Video and Audio)

APPENDIX 2
Additional Details about Methods

EdChoice hosted two back-to-back convenings in Fort Myers, Florida. Each convening spanned three days and two nights during the week of April 23–27, 2018. Each of the four affinity groups included six participants. However, there were a few exceptions when affinity groups were smaller because of last-minute absences. EdChoice staff led recruitment of the participants. Pluralism served as a guiding light for this project. We wanted to bring many different voices into the conversation and talk about the different approaches that can inform our policy choices and entrepreneurial decisions going forward regarding accountability in K–12 education.

We divided participants into one of four "affinity groups:"

Engaged Outsiders—Current non-educators, but work or volunteer in education-related organizations. They tend to have worn multiple professional hats in their established careers. Most are advanced in their careers and have achieved executive-levels positions. Some have launched their own education technology or service-providing companies.

Practitioners—Direct experience in schools, districts, or schooling organizations across various sectors, including public schools districts and district schools; public charter schools; private schools; and blended schools. Some are school leaders, teachers, school board members, or current and former superintendents.

Policy Advocates—Recognized education policy experts, working in educational advocacy organizations, state think tanks, statewide membership associations, or national public policy organizations.

Researchers—Education researchers who have specialized in a variety of fields including accountability in K–12 education. They are based at universities, state or national think tanks, or other nonprofit organizations that conduct education research and analysis and advance thought leadership.

In total, there were 47 participants from 20 states and D.C. Demographically, there were 24 men and 23 women. Ten identified as racial or ethnic minorities (7 African-American, 2 Hispanic, 1 Asian-American).

We pursued a full agenda. The total time of the substantive programming exceeded eight hours in either small group or large group sessions, not including breaks or meals. On the first day, we launched the program with an historic overview of the accountability movement in K–12 education. The rationale was to provide a shared knowledge base for all participants. Interpretation was minimal. The opening presentation emphasized history: timeline, people, events, etc. Hanover then facilitated three 75-minute discussion sessions for each of the four affinity groups.

Hanover's overarching rationale for this discussion facilitation design was that people tend to think "in the now." So the design takes participants on a pathway to develop ideas and opinions on this project's fundamental research questions: How did we get to where we are now regarding accountability in K-12 education? Where are we now? Where should we go?

Hanover used a focus group method because it is the optimal way to conduct an interactive discussion with participants of similar backgrounds or characteristics. Hanover recommended that the small affinity

---

[1] Insights Association (n.d.), Focus Groups [web page], accessed February 23, 2018, retrieved from http://www.marketingresearch.org/focus-groups

groups consist of people with similar roles. An effectively administered group stimulates interaction among participants and encourage them to form more developed opinions of key issues. Extant literature shows that group homogeneity maximizes disclosure and honesty among group participants. Participants should be similar in at least one clear way. Therefore, each small group consisted exclusively of one of the four affinity types: Engaged Outsiders, Practitioners, Policy Advocates, and Researchers. Small group composition remained constant throughout all small group sessions on the first day to allow for extensive, creative development of the group's insights.

Each small group discussion included a brief session introduction, primary questions, and exit questions (drafted by EdChoice). Hanover moderators led each small group and were responsible for introducing the topic and guiding the participants through the discussion.

The afternoon of the first day included a simulation session where each of the small groups had the opportunity to construct an accountability system for a hypothetical Midwestern state. That simulation was followed by a 60-minute large group session, where a panel composed of one moderator and one delegate from each small group presented their group's findings to all participants.

At the start of the second day, Hanover presented their initial findings about common themes and unresolved differences within affinity groups and across affinity groups for a given convening. An hour-long moderated discussion followed that presentation to allow participants to share their reactions, impressions, or to further elaborate on points made in the small group sessions or large group presentations. That synthesizing presentation led into the final large-group discussion on the future of accountability in K-12 education.

Following Hanover's advice, EdChoice staff encouraged participants to interact with others who were not in the same affinity group during session breaks and meals on the first day to allow for the informal sharing of ideas between participant groups and to avoid small group fatigue. Hanover moderators met during session breaks and meals to compare findings from respective small group sessions. These between-session moderator meetings were critical to informally capturing the developing trajectory of each small group discussion.

Questions about confidentiality are always present ahead of a data collection effort like this one. At the beginning of each small group discussion, Hanover informed participants that we were recording each session (via audio/video); how Hanover would share these recordings with EdChoice; and how Hanover would dispose of the recordings; how data collected during these discussions will be used, and by whom; and how each participant (based on consent) would be disclosed in a publication.

The authors and Hanover collaborated on a discussion guide with questions to serve as a general framework for the small group moderators. (see Appendix 2) While these questions provided a structure, a focus group method lends itself to free-flowing discussion. Participant comments, to some degree, frequently affect the direction of conversation and solicit the sharing of others. Consequently, the questions listed in Appendix 3 should not be viewed as exhaustive, nor a rendering of the actual small-group discussions.

---

[2] Eliot and Associates (2005), Guidelines for Conducting a Focus Group, retrieved from Data Innovation Project website: https://datainnovationproject.org/wp-content/uploads/2017/04/4_How_to_Conduct_a_Focus_Group-2-1.pdf
[3] Ibid; The National Institute for Urban School Improvement (2005), *Conducting Focus Groups to Develop a Comprehensive School Portrait*, p. 3, retrieved from http://www.niusileadscape.org/docs/FINAL_PRODUCTS/LearningCarousel/conductingfocusgroupsfinal.pdf; Insights Association (n.d.), Focus Groups [web page], accessed February 23, 2018, retrieved from http://www.marketingresearch.org/focus-groups
[4] Ibid., p. 2

APPENDIX 3
Discussion Guide for Sessions

These questions combine to serve as a general framework that the small group facilitators will adjust as needed. While these questions provide a structure, small group methodology lends itself to free-flowing discussion. Ideally, participant comments will, to some degree, affect the direction of conversation and solicit the sharing of others. Consequently, the questions listed in this section should not be viewed as a finite and inflexible agenda.

Small Group Session 1
Lessons Learned (75 Minutes)

General Questions and Time Limit	Follow-Up Questions	Objective
How would you describe the lessons we can(have) learn from the modern accountability movement, post-2001? **(20 minutes)**	• How did we get here? • What were the most significant moments in the development of the current state of accountability? • What are some of the less obvious influences that led us to the current state of accountability?	This question functions as an ice breaker, warming up participants to one another, to talking as a group, and to thinking about trajectories of the past.
What are the biggest triumphs of the modern accountability movement, post-2001? **(20 minutes)**	• What good has come from the movement? • What lessons can be learned from the history of accountability about how to create more good?	This question functions as an ice breaker, warming up participants to one another, to talking as a group, and to thinking about trajectories of the past.
What are the biggest mistakes of the modern accountability movement, post-2001? **(20 minutes)**	• Describe for me the outcomes of those mistakes. • What lessons can be learned from the history of account ability about how to avoid more mistakes?	This question functions as an ice breaker, warming up participants to one another, to talking as a group, and to thinking about trajectories of the past.

Of all the things we have discussed in this session, what do you think is the most important?

Final thoughts?
Brief preview of next discussion
Selections of small group delegate

(15 minutes)

Small Group Session 2
Accountability Today (75 Minutes)

General Questions and Time Limit	Follow-Up Questions	Objective
What does it mean to have accurate, actionable information on school performance? **(20 minutes)**	• How would you describe the current discourse surrounding accountability in K–12 education? • How is "accountability" understood differently by different K–12 stake holders? • How do different K–12 stake holders talk to one another about accountability?	This question aims to uncover shared beliefs and differences of opinion about what accountability means and how K–12 stakeholders talk about it.
What changes are currently being instituted in schools because of accountability? **(20 minutes)**	• How are curricula changing? • How are teacher behaviors changing? • How are student behaviors changing? • How would you describe the direction of these changes?	This question aims to gather knowledge from participants about recent consequences of accountability from their on-the-ground experiences.
What are the biggest problems with the current accountability systems? **(20 minutes)**	• How are these problems related to one another? • How might these problems be tackled with single solutions?	This question functions as an ice breaker, warming up participants to one another, to talking as a group, and to thinking about trajectories of the past.

Of all the things we have discussed in this session, what do you think is the most important?

Final thoughts?
Brief preview of next discussion

(15 minutes)

General Questions and Time Limit	Follow-up Questions

You have been appointed by the Governor of a Midwestern state to draft the state's new accountability program. The state has traditional public schools, magnet schools, charter schools, and education savings account programs.
 	
What are the key design elements of your accountability program?
 	
(30 minutes)
	• What information and data would you collect for policymaking and decisions?
	
	   - What are the pros and cons of the various sources of information/data?
	
	• How would you use that information/data?
	
	   - What does transparency look like?

Who would you include to consult and advise on policy and decisions? **(15 minutes)**	• Why are those people, organizations important to you? • To what extent do you value expert recommendations, parental decisions, other stakeholder considerations?

Small group preparation of delegate for large group discussion

Of all the things we have discussed in this session, what do you think is the most important? Final thoughts?

(15 minutes)

General Questions and Time Limit	Follow-Up Questions	Objective
"What are the key values, principles, and design features of tomorrow's K–12 accountability system?"	• How can we learn from past lessons to avoid future mistakes? • How can we learn from past lessons to avoid future unintended consequences?	This question functions to warm up participants to the session topic and spark their thinking about trajectories of the future.
How does the ideal system of accountability integrate schools of choice?	• How does the ideal system of accountability integrate schools with unique missions?	This question aims to include schools of choice and schools with unique missions in the discussion of future accountability systems.
What does it mean to have actionable information in a system of accountability?	• Where would this information be found? • Who should be responsible for collecting this information? • What about for deciding how it is used?	This question functions as an ice breaker, warming up participants to one another, to talking as a group, and to thinking about trajectories of the past.
How do we measure success in K–12 accountability?	• How do those measures mitigate challenges we have encountered in the past? • What do we need to institute those measures?	This question aims to close the focus group series by exploring how to measure the success of a future accountability system.

Of all the things we have discussed in this session, what do you think is the most important?

Final thoughts?
Brief preview of next discussion

(15 minutes)

NOTES

1 George W. Bush (2002, September 4), *Remarks on Implementation of the No Child Left Behind Act of 2001* [Speech transcript], retrieved from The American Presidency Project website: http://www.presidency.ucsb.edu/ws/?pid=73078

2 Barack Obama (2012, February 9), *Remarks by the President on No Child Left Behind Flexibility* [Speech transcript], retrieved from https://obamawhitehouse.archives.gov/the-press-office/2012/02/09/remarks-president-no-child-left-behind-flexibility

3 National Center for Education Statistics (1993), *120 Years of American Education:* A Statistical Portrait, retrieved from https://nces.ed.gov/pubs93/93442.pdf

4 Hugh Davis Graham (2011), *The Uncertain Triumph: Federal Education Policy in the Kennedy and Johnson Years*

5 United States National Commission on Excellence in Education (1983). *A Nation At Risk: The Imperative for Educational Reform: A Report to the Nation and the Secretary of Education*, United States Department of Education, p. 28, retrieved from https://eric.ed.gov/?id=ED226006

6 Martin Carnoy and Susan Loeb (2002), Does External Accountability Affect Student Outcomes? A Cross-State Analysis, *Education Evaluation and Policy Analysis*, *24*(4), pp. 305–11, https://dx.doi.org/10.3102%2F01623737024004305

7 David Figlio and Susanna Loeb (2011), School Accountability, in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann (eds.), Handbooks in Economics (Volume 3), p. 384, retrieved from https://cepa.stanford.edu/sites/default/files/Accountability_Handbook.pdf

8 The Education Trust–West, Equity & Accountability: What You Need to Know [Blog post], retrieved from https://west.edtrust.org/equity-accountability-what-you-need-to-know/

9 Accountable (n.d.), In *Merriam-Webster's online dictionary*, retrieved from https://www.merriam-webster.com/dictionary/accountable?

10 Magnet Schools of America, Key Facts About Magnet Schools [web page], accessed August 16, 2018, retrieved from http://magnet.edu/govt-relations/grassroots-action-center/key-facts-about-magnet-schools

11 National Alliance for Public Charter Schools (2017), *Estimated Charter Public School Enrollment, 2016–17*, retrieved from http://www.publiccharters.org/sites/default/files/migrated/wp-content/uploads/2017/01/EER_Report_V5.pdf

12 EdChoice (2018), *The ABCs of School Choice: The Comprehensive Guide to Every Private School Choice Program in America, 2018 Edition*, retrieved from https://www.edchoice.org/wp-content/uploads/2018/01/ABCs-of-School-Choice-2018-Edition-1.pdf

13 Greg Forster (2015, June 29), Choice is the Real Accountability, *Oklahoma Council of Public Affairs*, retrieved from http://www.ocpathink.org/post/choice-is-the-real-accountability?print

14 Phi Delta Kappan (2017), *The 49th Annual PDK Poll of the Public's Attitudes Toward the Public Schools*, http://journals.sagepub.com/doi/pdf/10.1177/0031721717728274

15 James P. Kelly, III and Benjamin Scafidi (2013), *More Than Scores: An Analysis of Why and How Parents Choose Private Schools*, p. 6, retrieved from EdChoice website: http://www.edchoice.org/wp-content/uploads/2015/07/More-Than-Scores.pdf#page=6

16 We have lightly edited some participant quotes from transcripts for readability, such as deleting filler words for example. The bracketed text within quotes indicates our editing.

17 Donald T. Campbell (1979), Assessing the impact of planned social change, Evaluation Program and Planning, 2(1), pp. 67–90, https://dx.doi.org/10.1016/0149-7189(79)90048-X

18 UChicago Impact, The 5 Essentials System [web page], accessed August 16, 2018, retrieved from https://uchicagoimpact.org/our-offerings/5essentials

19 NICE Satmetrix, What is Net Promoter? A Trusted Anchor for your Customer Experience Management Program, accessed August 16, 2018, retrieved from https://www.netpromoter.com/know

20 Paul DiPerna and Andrew D. Catt (2016), *2016 Schooling in America Survey: Public Opinion on K–12 Education and School Choice*, Polling Paper 28, p. 20, retrieved from EdChoice website: https://www.edchoice.org/wp-content/uploads/2016/11/2016-10_SIA-Poll-Update.pdf; Phi Delta Kappan (2017), The 49th Annual PDK Poll of the Public's Attitudes Toward the Public Schools, http://journals.sagepub.com/doi/pdf/10.1177/0031721717728274

21 Kate McGee (2017, November 28), What Really Happened at the School Where Every Graduate Got Into College [Radio transcript], *National Public Radio*, retrieved from https://www.npr.org/sections/ed/2017/11/28/564054556/what-really-happened-at-the-school-where-every-senior-got-into-college

ABOUT THE AUTHORS

Michael Q. McShane

Michael Q. McShane is director of national research at EdChoice. He is the editor of New and Better Schools (Rowman and Littlefield, 2015), the author of Education and Opportunity (AEI Press, 2014), and coeditor of Failure Up Close: What Happens, Why it Happens, and What We Can Learn from It (Rowman and Littlefield, 2018), Educational Entrepreneurship Today (Harvard Education Press, 2016), Teacher Quality 2.0 (Harvard Education Press, 2014), and Common Core Meets Education Reform (Teachers College Press, 2013). His analyses and commentary have been published widely in the media, including in the Huffington Post, National Affairs, USA Today, and The Washington Post. He has also been featured in education-specific outlets such as Teachers College Commentary, Education Week, Phi Delta Kappan, and Education Next. In addition to authoring numerous white papers, McShane has had academic work published in Education Finance and Policy and the Journal of School Choice. A former high school teacher, he earned a Ph.D. in education policy from the University of Arkansas, an M.Ed. from the University of Notre Dame, and a B.A. in English from St. Louis University. He is an adjunct fellow in education policy studies at the American Enterprise Institute and a Senior Fellow at the Show-Me Institute.

Paul DiPerna

Paul DiPerna is vice president of research and innovation for EdChoice. He joined the organization in 2006. Paul's research interests include surveys and polling on K–12 education and school choice reforms. He oversees the research projects either produced or commissioned by the organization. EdChoice has published more than 95 reports, papers, and briefs during his tenure leading the research program. Paul has traveled to 31 states for his work. He presents survey research findings and discusses school choice politics and policies with audiences, including public officials, policy professionals, academics, and advocates. Paul's professional memberships and activities include participation in the American Association for Public Opinion Research (AAPOR) and Association for Education Finance and Policy (AEFP). Previously, Paul served as the assistant director for the Brown Center on Education Policy at the Brookings Institution. He was a research analyst for the first five issues of the Brown Center Report on American Education (2000–2004). He also managed and coordinated the activities of the National Working Commission on Choice in K–12 Education (2001–2005). A native of Pittsburgh, Paul earned an M.A. in political science from the University of Illinois (2000) and B.A. from the University of Dayton (1996). Paul currently lives in Zionsville, Indiana, with his wife and two daughters.

ACKNOWLEDGMENTS

COMMITMENT TO METHODS & TRANSPARENCY

EdChoice is committed to research that adheres to high scientific standards, and matters of methodology and transparency are taken seriously at all levels of our organization. We are dedicated to providing high-quality information in a transparent and efficient manner.

The American Association for Public Opinion Research (AAPOR) welcomed EdChoice to its AAPOR Transparency Initiative (TI) in September of 2015. The TI is designed to acknowledge those organizations that pledge to practice transparency in their reporting of survey-based research findings and abide by AAPOR's disclosure standards as stated in the Code of Professional Ethics and Practices.

All individuals have opinions, and many organizations (like our own) have specific missions or philosophical orientations. Scientific methods, if used correctly and followed closely in well-designed studies, should neutralize these opinions and orientations. Research rules and methods minimize bias. We believe rigorous procedural rules of science prevent a researcher's motives, and an organization's particular orientation, from pre-determining results.

If research adheres to proper scientific and methodological standards, its findings can be relied upon no matter who has conducted it. If rules and methods are neither specified nor followed, then the biases of the researcher or an organization may become relevant, because a lack of rigor opens the door for those biases to affect the results.

The contents of this publication are intended to provide empirical information and should not be construed as lobbying for any position related to any legislation.

The authors welcome any and all questions related to methods and findings.

ed CHOICE

111 MONUMENT CIRCLE
SUITE 2650
INDIANAPOLIS, IN 46204
317 681 0745

EDCHOICE.ORG