Multimetric Accountability Systems

A Next-Generation Vision of Student Success

MELISSA MELLOR // DAVID GRIFFITH

Multimetric Accountability Systems

TABLE OF CONTENTS

ASCD
LEARN. TEACH. LEAD.

WWW.ASCD.ORG

1703 North Beauregard St.
Alexandria, VA 22311–1714
1-703-578-9600 or
1-800-933-2723

INTRODUCTION

When the No Child Left Behind (NCLB) Act became law well over a decade ago, one of its primary goals was to draw attention to and try to address persistent achievement gaps among the nation's students. It did this by ushering in an unprecedented level of federal accountability for student academic achievement. NCLB requires every public school to report disaggregated student achievement results on English language arts and mathematics standardized assessments. If schools fail to make sufficient annual gains in test scores—both overall and for specific subgroups of students—they face increasingly severe sanctions, from restructuring to state takeover and even closure. NCLB has been widely praised for highlighting the nation's performance disparities among different groups of students, but its test-based and punitive accountability measures have been broadly criticized for their unintended consequences of narrowing the curriculum, encouraging too much focus on test preparation, and de-emphasizing important factors related to student well-being and achievement. The law's approach has clearly reached its limits, and it's time for a more meaningful next phase of school accountability.

There's growing consensus that this next phase must align with better drivers of school improvement and student success. Michael Fullan defines the appropriate drivers as those that promote capacity building, foster the intrinsic motivation of teachers and students, encourage collaboration, and build coherence across systems (2011). At the same time, there's been a push to define student success more broadly. David Conley's research has identified four keys to college and career readiness: cognitive strategies, content knowledge, learning skills and techniques, and the ability to transition to life beyond high school (n.d.). Still, others would add citizenship, social-emotional learning skills, or character education to their definitions of student success (Michelman, 2015).

Based on the association's whole child approach, ASCD has been calling for more meaningful accountability systems that promote continuous support and improvement and align with the broader outcomes we collectively want for our students. In particular, such systems should incorporate a variety of measures that more fully reflect a comprehensive definition of student success, accurately measure student learning, and systematically track educators' efforts to engage and support learners.

It's worth noting that ASCD's definition of multiple measures extends beyond multiple tests (even if they are different types of tests or tests in other subject

areas) and multiple ways of analyzing test scores (e.g., absolute performance, achievement gaps, and growth). For accountability systems to both capture and incentivize the full breadth of learning and support in schools, they must also incorporate measures beyond standardized tests.

Most educators and policymakers now recognize the value of this multiple measures approach—what we at ASCD call multimetric accountability. Yet broad questions remain about what measures are reliable and valid for use as accountability metrics. At the same time, local systems have specific questions about which measures to use, what combination of measures constitutes an appropriate mix, how the measures should be weighted (if they are weighted at all), and how to effectively communicate about the measures and progress toward meeting them. To begin to answer these questions, this paper provides five real-world examples of accountability systems from Tacoma, Washington; California's CORE districts; New Hampshire's PACE districts; Kentucky; and Alberta, Canada.

ASCD's analysis of these systems reveals that in response to the first question, there is no shortage of reliable and valid measures that can be used for school accountability purposes. As for the subsequent questions, there is no right answer; instead, each district, state, or province needs to collaboratively determine what is best for its students. ASCD's intent is that these examples will highlight the multimetric accountability work that is currently being done, enhance the understanding of what multimetric accountability systems can look like and how they work, and share the initial benefits and challenges of this approach.

This paper is by no means an exhaustive analysis of multimetric accountability systems or an endorsement of particular strategies or approaches over others not referenced. For example, several states have begun using college- and career-readiness indicators in addition to the assessment data and graduation rates already in their accountability models. And many districts have added their own accountability measures to enhance the metrics tracked by their states. In addition, the paper focuses on the multiple measures component of accountability and does not delve deeply into other aspects such as the identification of low-performing schools or school improvement strategies. But the following five examples represent some of the more comprehensive and innovative multimetric accountability systems—at the district, state, and provincial levels—that we've been able to find.

Taken together, they reveal that there's no single pathway for incorporating multiple measures into accountability and that there's a plethora of measures that can

be included. Despite these differences, the examples illuminate some common principles and challenges when implementing a multimetric accountability system. Most important, they are all premised by the shared belief that any accountability system must be rooted in a vision of student success that is established well before specific measures are identified. After all, the power of accountability is to transparently indicate how well schools and communities are preparing students for such success and informing continuous improvement toward that goal. That's impossible if the accountability model isn't created with that end in mind.

TACOMA PUBLIC SCHOOLS

Holding Schools and the Community Accountable for the Whole Child

Tacoma Public Schools is the third largest school district in Washington State, and it has become known for its commitment to improvement as well as its inventive ideas and practices. This spirit of innovation is evident in the district's homegrown accountability system, which holds schools accountable not only for student performance on tests but also for broader efforts to improve their students' academic achievement, engage families and communities in learning, and maintain safe and healthy learning environments.

System Profile

Superintendent: Carla Santorno
www.tacoma.k12.wa.us
Student Enrollment: 29,348

American Indian: 1%
Asian: 10%
Asian/Pacific Islander: 13%
African American: 20%
Hispanic: 17%
Multiracial: 4%
Native Hawaiian/Other Pacific Islander: 2%
White: 45%

Free or Reduced-Price Meal Eligibility: 63%
Special Education: 14%
Transitional Bilingual: 9%

These efforts were prompted by citywide dissatisfaction with the district's high school graduation rate and a collective call for change. In 2010, Tacoma's graduation rate dropped to an all-time low of 55 percent, and every one of its high schools was deemed a "dropout factory" by a national researcher and the media. In response, district officials and Tacoma community members began working together on a new approach to accountability that more comprehensively supported and prepared students. The resulting system strategically aligned with the districts' overall purpose of supporting the whole child and reflected the belief that success for Tacoma students means more than a one-time score on a standardized test.

Tacoma started with its four main goals for its students: academic excellence and the elimination of disparities among student groups; partnerships that engage parents, community, and staff; early academic success; and safe learning

environments. Then it began identifying the best measures to show how well schools and the district are doing in each of those areas. Through an intensive feedback process involving a wide variety of stakeholders, from educators to community partners, the district eventually identified more than 40 indicators that aligned with the four goals. The resulting indicators range from typical accountability metrics such as student assessment results and graduation rates to less common measures such as student participation in extracurricular activities and rigorous coursework; school climate survey participation and results; preschool enrollment rates; and even the proportion of graduates with verified acceptance letters from a postsecondary pathway (e.g., community colleges, universities, the military, and apprentice programs). For a full list of Tacoma's accountability measures, go to www.tacoma.k12.wa.us/benchmarks/Pages/default.aspx.

"Tacoma's journey in creating benchmarks was critical to building the community's trust in the system."

The process for designing its school accountability system may seem straightforward in retrospect, but Joshua Garcia, Tacoma's deputy superintendent, cautions that getting an entire community to define success for every child in a measureable way is extremely difficult. He emphasizes, however, that the challenging collaborative work was worth it, and "Tacoma's journey in creating benchmarks was critical to building the community's trust in the system." It's also worth noting that the journey to create those benchmarks is not over. Tacoma continues to refine its measures and tweak the language it uses to describe results so that the district is as clear as possible about student progress.

All of the accountability indicators reflect student outcomes as opposed to adult measures like teacher evaluation results—a decision that Garcia believes facilitated the community's approval of the system. In addition, school and district performance on the metrics are reported individually and are not rolled up into a single score. "We want to be extremely transparent with our community about where we are with every one of our agreed benchmarks," says Garcia. "You can't have a single grade and truly get clear about whether kids are healthy, safe, engaged, supported, and challenged based on our local definition." School performance on the measures determines not only interventions for schools that are struggling in certain areas but also increased autonomy and flexibility for those schools that are performing well. Tacoma's system meets all state and federal

assessment and accountability requirements, but Garcia explains that it does so within the context of local values and global competitiveness.

Accountability in Service of Continuous Improvement

The true value of Tacoma's accountability approach is that it drives continuous improvement across the district. A key step that facilitated this was presenting the community-recommended metrics to the board for approval. The board's adoption of the metrics as part of its five-year strategic plan created clarity and consensus on goals, which in turn drives crucial decision making related to resource allocation, district policies, professional development, and strategic partnerships. Take, for example, the district's goals for improving student participation in rigorous courses and boosting the proportion of students who have been accepted into postsecondary institutions. Tacoma identified the barriers—such as qualification criteria and financial challenges—that were preventing some students from succeeding with those goals. Then it "relentlessly worked to remove those barriers," according to Garcia. It adopted a policy that requires students to "opt out"—instead of "opt in"—to college level courses and allows all students to take Advanced Placement and International Baccalaureate exams for free and all high school seniors to take the SAT for free. It also increased communication with families about student performance and worked with partners to establish multiple scholarship opportunities.

Three years into this new approach, Tacoma is already seeing results. The district's graduation rate for the class of 2014 has soared to 78 percent—the highest it's ever been since the state began officially tracking the statistic more than a decade ago. In addition, Tacoma's graduation rates for every racial demographic have improved; for example, the graduation rate for black students climbed from 59 percent in 2012 to 74 percent in 2014. Tacoma's preschool enrollment rate has also increased. This year, 1,660 students attended preschool compared to 1,190 in 2010. Washington State recently recognized a dozen Tacoma schools for exceptional academic performance in 2014. Previously, no more than five Tacoma schools in a given year had been recognized.

Change hasn't been easy, and district officials don't believe that Tacoma has developed the perfect accountability system. But after witnessing the gains that have been made, Garcia thinks Tacoma has created something that's far better than the status quo. "We believe passionately that every community should use a multiple-measure approach that represents their definition of what a whole child education includes," he says.

CORE DISTRICTS

Valuing and Incentivizing Efforts to Prepare Students for Success in Life

System Profile

Executive Director: Rick Miller
www.coredistricts.org
Participating Districts: Clovis, Fresno, Garden Grove, Long Beach, Los Angeles, Oakland, Sacramento, San Francisco, Sanger, and Santa Ana Unified
Student Enrollment: 1,108,659
African American: 8%
Asian: 8%
Hispanic: 62%
White: 11%

Free or Reduced-Price Meal Eligibility: 74%
English Language Learner: 26%

In 2013, the CORE districts, a collaboration of 10 California school districts representing more than one million students—about 18 percent of the state's students—received the first and only district-level waiver from NCLB's accountability provisions. The CORE districts originally formed in the aftermath of a fruitful, although ultimately unsuccessful, cross-district effort to apply for a statewide Race to the Top grant. Even though the districts lost the federal grant competition, they recognized the value of collaborating and decided to form an organization and engage in peer-to-peer learning with some seed funding from a few California foundations. When California decided not to seek a statewide NCLB waiver, the CORE districts were given an unprecedented opportunity to apply, and they were granted a waiver.

The waiver exempted the districts from California's version of the federal NCLB accountability system, requiring them to build their own system in its place. The CORE districts began developing their new school accountability approach based on three key principles:

- The accountability system would facilitate work that educators value and were already doing—work that raises achievement for all students.
- It would promote a higher level of mutual responsibility.
- It would be driven by equity and a focus on eliminating disparity and disproportionality.

Up to this point, schools in these districts had been subject to three different accountability models or reporting mechanisms: the federal adequate yearly progress model required under NCLB, the state's Academic Performance Index (API), and the districts' own data dashboards. While the first two models afforded narrow views of the health of schools and students, the districts' dashboards were much richer and more nuanced groupings of local data that they valued. That's

why the CORE districts turned to their respective dashboards as a starting point for developing a new, collaborative approach to accountability that prioritized the right drivers. CORE wanted its new accountability system, called the School Quality Improvement System, to use metrics that would ultimately help students be prepared for college and careers.

Several of the districts were deeply involved in improving school culture and climate and building students' social-emotional skills. During initial conversations, there was significant debate about whether to make these components part of CORE's shared accountability framework. But once the districts embraced the philosophy that their new accountability approach would reflect what matters, the question shifted from *whether* to incorporate those components to *how* best to incorporate them.

The School Quality Improvement System has two main domains: an academic domain and a social-emotional/culture-climate domain. Within the academic domain, the accountability metrics include assessment results (both current performance and growth), a high school readiness indicator that captures the percentage of 8th graders who meet a set of criteria that predict they are likely to graduate high school on time, and graduation rates. The social-emotional and culture-climate domain includes a chronic absenteeism rate, suspension and expulsion rates, school climate survey responses, and student self-responses on surveys that ask about their social and emotional skills (see Table 1).

Table 1: CORE Districts' School Quality Improvement System Accountability Index

Domains	**Metrics**
Academic (60% of the accountability index score)	English and math current performance on Smarter Balanced Assessments
	Individual student growth on Smarter Balanced Assessments
	High school readiness rate
	Graduation rates (four-year, five-year, and six-year)
Social-Emotional and Culture-Climate (40% of the accountability index score)	Chronic absenteeism rate
	Student, staff, and parent responses on school culture-climate surveys
	Suspension/expulsion rates
	Student self-reports of social-emotional skills
	English learner redesignation rate

Incorporating SEL Skills

CORE recognizes that its reliance on survey data, particularly with regard to students' social-emotional skills, pushes into unchartered territory; in fact, the CORE districts are participating in the largest survey of students' social-emotional skills ever conducted. The survey focuses on four specific social-emotional learning (SEL) competencies—growth mind-set, self-efficacy, self-management, and social awareness—that were selected based on research about the importance, measurability, and actionability of each as well as the experiences of educators within the participating districts. A pilot study of the survey instrument revealed that while there are limitations to using student-reported data, the results met reliability and validity measurement criteria. CORE is currently working with a number of research organizations and SEL thought leaders to explore SEL-related performance tasks and other "next generation" measures, such as games and situational judgment tasks. Ultimately, the districts would prefer to use performance assessments to determine students' SEL skills and school quality reviews to gain a sense of how schools' SEL commitments are translating at the classroom level, says CORE's executive director Rick Miller. But, for now, the surveys are the best available instrument.

All CORE district schools will receive a single accountability index score, with the academic domain measures accounting for 60 percent of the score and the SEL and school climate metrics constituting the remaining 40 percent. However, Miller emphasizes that CORE is most interested in school performance on the individual metrics. Plans are in the works for a district dashboard that will be rolled out in the fall of 2015 where schools and communities will be able to view performance in each area of the accountability system.

CORE is committed to using its accountability determinations to strategically build schools' capacity—not to sanction them. For example, its school-pairings process provides support and technical assistance to schools that need it from partner teachers and school leaders that are achieving success with similar students in other schools.

CORE's leaders acknowledge that the full results of their innovative approach remain to be seen; its complete slate of metrics will not be used to publically evaluate schools and inform their improvement until the 2015–16 school year. However, they believe the development work alone has proven fruitful. Drilling down into how they are defining each of the measures and engaging in some "productive conflict" has clarified priorities and given the districts a better idea of what success looks like. Noah Bookman, CORE's chief accountability officer,

explains, "Just seeing the range of metrics in the accountability system has energized and inspired CORE's schools. It's validated that all of the components are things they are supposed to attend to."

Miller adds, "Past experience has clearly shown us that accountability systems based only on academic achievement results are problematic. We're hoping that our work provides a useful test case and an example of what's possible."

NEW HAMPSHIRE PACE PILOT

Shifting from State-Driven Accountability to Mutual Responsibility

This past spring, the U.S. Department of Education provided New Hampshire with an NCLB waiver that allows four of the state's districts to pilot a first-of-its-kind locally managed performance assessment system for state and federal accountability purposes. The New Hampshire Performance Assessment for Competency Education (PACE) pilot permits the districts to give the Smarter Balanced statewide assessment once during each grade span (elementary, middle, and high school) instead of requiring it annually in grades 3–8 and once in high school. For the remaining grades, the districts administer locally developed performance assessments in English language arts, math, and science. Much attention has been paid to how the two-year PACE pilot (which is being implemented during the 2014–15 and 2015–16 school years) is reducing the reliance on standardized testing by essentially aggregating federal and state assessment and accountability requirements into one system. But New Hampshire's deputy commissioner of education, Paul Leather, also emphasizes that the new approach is more relevant to teaching and learning practices than the current accountability model and more deeply invests educators in the work. During the 2015–16 school year, four additional districts will be joining the pilot.

System Profile

State Commissioner: Virginia M. Barry
http://education.nh.gov/assessment-systems/pace.htm
Participating Districts: Sanborn Regional, Rochester, Epping, and Souhegan
Student Enrollment: 8,014
White: 92%
Free or Reduced-Price Meal Eligibility: 33%

A confluence of factors have led New Hampshire to this point. Since the late 1990s, the state has embraced competency-based education. In 2005, New Hampshire required all of its high schools to move from a traditional Carnegie Unit system that awards students credit for learning based on time spent in a course to a competency-based system that recognizes true mastery of learning. For example, if students demonstrate the knowledge and skills that will be

learned in a particular course, they can receive credit for that course and enroll in a more advanced level of the subject or a different course entirely. Outside-the-classroom learning experiences such as internships at a local company can also count toward a student's graduation credits. This shift to a competency-based system necessitated innovations in assessment and grading practices that provide students with multiple authentic opportunities to demonstrate their knowledge and skills at various points throughout the year.

In addition, New Hampshire is part of the Council of Chief State School Officers' Innovation Lab Network (ILN)—a group of states that is implementing student-centered approaches to learning. The ILN has explored how building more locally driven systems of accountability can align with efforts to promote personalized learning experiences and performance-based assessments. These and other factors prompted New Hampshire to move away from a large-scale state assessment and accountability model to a mutual responsibility system in which local districts are taking more ownership for defining and measuring student learning and growth.

The four districts (Sanborn Regional, Rochester, Epping, and Souhegan) that are participating in the PACE pilot this year represent approximately 4 percent of the state's student population. Rochester is a relatively large school district with 11 schools; the others are smaller, like most of the state's districts. Souhegan represents an upper-middle-class collection of communities, and the other districts are more typical of the populations served by districts across the state. The four districts had to meet several prerequisites in order to take part in the pilot. They had to adopt the state graduation competencies in English language arts, math, and science; demonstrate that their teachers were using a competency-based approach to education in personalized learning environments; participate in state-sponsored quality performance assessment team training; develop their own performance tasks; and generally demonstrate the necessary capacity to implement the pilot with fidelity. For this reason, none of the participating districts currently has "priority" schools (schools that are among the lowest performing in the state). The districts have also committed to administering at least one performance task in each grade and in each content area that is common across all of the districts.

Developing Effective and Comparable Performance Assessments

As part of their participation in the pilot, the districts will engage in a peer review of their local competencies and assessment system plans. The peer review will determine whether the districts' performance expectations for students are

appropriately rigorous and thus comparable within schools, among schools, and among districts. The review will also ensure that districts are using a coherent trajectory of content competencies leading to the state model graduation competencies and that the local PACE assessments are clearly measuring student mastery of those competencies. The PACE districts' annual accountability determinations will be based on an aggregation of local and state performance assessment results as well as Smarter Balanced scores. The exact weights of the various assessments and other factors like graduation rates in the accountability calculation will be determined this August.

The state will review the districts' initial achievement and growth results, compare the results to the Smarter Balanced assessment results, and review the performance tasks for technical quality in terms of alignment, reliability, fairness, and competency determinations. New Hampshire is also providing significant support to the participating districts through professional development institutes that focus on assessment literacy, competency-based instruction, data analysis, and more. In addition, regional task validation and calibration scoring sessions will help districts refine their assessment tasks and build inter-rater reliability and consistency in scoring. The state is maintaining a performance task bank that accepts tasks from PACE districts and other New Hampshire schools and vets the tasks on an ongoing basis. Currently, about 120 tasks are in the bank, and state officials expect that number to double, if not quadruple, in one year.

New Hampshire is continually thinking about how to build on this work. During the 2015–16 school year, the state plans to expand the assessment window to allow the districts and schools to test students when they are ready and not according to a predetermined schedule—a decision in keeping with a competency-based approach to education. Eventually, the state would like districts to administer performance tasks in other subjects. In fact, New Hampshire has completed model competencies for the arts and will turn its attention to social studies next. Although state officials decided to initially focus the PACE pilot on English language arts, math, and science so as not to overwhelm the participants, Leather believes the districts and schools will consider adding other subjects over time. "If you make assessment and accountability part of meaningful work," says Leather, "it will change educators' perspective, and they will see those components as essential and not just another thing to do."

It should be noted that the current performance assessments also evaluate students' work study practices, such as communication skills, creativity, ability to collaborate, and self-direction. However, the work study practices results are not

being used for accountability purposes. Instead, schools can choose to use them as part of local grades or student learning objectives.

Ultimately, New Hampshire officials hope all of its districts will adopt the PACE model over the next 10 years. Leather shares the two key questions New Hampshire leaders must grapple with as they work to meet this goal: "How many districts can the state support?" and "How much support will the state receive in pursuing this approach?" In fact, state officials have had conversations with federal policymakers about the need to shift away from universal adherence to one accountability system to a staged approach where rolling cohorts of districts and schools adopt new accountability models as they demonstrate readiness. An overarching concern is boosting struggling schools' capacity to take on this meaningful but difficult work.

> ***"If you make assessment and accountability part of meaningful work, it will change educators' perspective, and they will see those components as essential and not just another thing to do."***

Although the accountability component of the PACE pilot is just beginning, Leather offers some key insight based on his years of experience in helping New Hampshire schools move to competency-based education. First, he contends that the state's approach is promoting much deeper learning than was possible before. Second, children are being provided with more entry points for learning. "This part is critical," says Leather. "Competency-based education can stop kids from falling by the wayside, and our data show that this is the case. New Hampshire was one of the first states to increase high school graduation rates and reduce dropout rates." Finally, Leather emphasizes that teacher effectiveness is crucial to the success and sustainability of New Hampshire's efforts: "We started with professional development to prepare educators to help their students reach deeper levels of learning and apply their skills in complex tasks. We did this several years before making competency-based education and performance assessments part of accountability."

KENTUCKY

Honoring a Well-Rounded Curriculum

In 2009, Kentucky Education Commissioner Terry Holliday began discussions with the Kentucky Board of Education and various education stakeholder groups about the broad concept of a new state accountability model. At the core of their conversation was the desire for a more balanced accountability system anchored in college and career readiness for all students. The state's leaders were looking to expand on NCLB's minimum accountability requirements. They believed that school accountability creates pressure points and that having too few pressure points can contribute to an overly narrow focus on certain education outcomes. In 2011, the leaders' vision for a more balanced approach became reality when Kentucky adopted its new Unbridled Learning accountability model.

System Profile

State Commissioner: Terry Holliday
www.education.ky.gov
K–12 Enrollment: 654,289

African American: 11%
Asian: 2%
Hispanic: 5%
Multiracial: 3%
White: 80%

Free or Reduced-Price Meal Eligibility: 58%
English Language Learner: 3%
Special Education: 13%

The Unbridled Learning accountability model has three main components: 1) next-generation learners, 2) next-generation instructional programs and support, and 3) next-generation professionals. Rhonda Sims, associate commissioner in the Kentucky Department of Education's Office of Assessment and Accountability, explains that the new model has been built to ensure that the state's students are leaving the K–12 education system ready for their next steps, whether that's further education or entry-level work.

Within the learners component, the accountability metrics include state assessment results (current performance, achievement gaps, and growth), college- and career-readiness rates, and high school graduation rates. It's worth noting that Kentucky is one of very few states to include assessment results in its accountability system that extend beyond English language arts, math, and science to also include results for social studies and writing. The college- and career-readiness rate at the middle school level is the proportion of students meeting benchmarks on the ACT EXPLORE assessment, which is administered in grade 8. At the high school level, there are multiple pathways for students to demonstrate their readiness for college and career, including meeting the state's college-readiness benchmarks on the ACT (which is taken by all 11th graders), passing a college placement test such as ACT's COMPASS, or attaining career-ready academic and

technical benchmarks on examinations or by receiving industry-recognized career certificates.

The system's instructional programs and support component includes program reviews of schools' arts and humanities, practical living and career studies, writing, early learning, and—for the first time starting with the 2015–16 school year—world language programs. The final component, professionals, will also come on board starting with the 2015–16 school year and will include the percentages of effective teachers and leaders as determined through Kentucky's statewide evaluation system. Up to now, the next-generation learners metrics make up 77 percent of a school's final accountability score, and the program reviews make up the remaining 23 percent. Once the professionals piece is worked in, the weights will be 70 percent, 20 percent, and 10 percent, respectively (see Table 2).

Table 2: Kentucky's Unbridled Learning Accountability Model

Components	Metrics
Next-Generation Learners (70% of the overall accountability score)	Current achievement on standardized tests (reading, math, science, social studies, and writing)
	Gap calculation (standardized test performance in all five subject areas of students belonging to groups that have historically had achievement gaps)
	Student growth in reading and math achievement on standardized tests
	College/career readiness rate
	Graduation rate
Next-Generation Instructional Programs and Support (20% of the overall accountability score)	Program reviews in arts and humanities, practical living and career studies, writing, early learning, and world languages
Next-Generation Professionals (10% of the overall accountability score)	Percentage of effective teachers
	Percentage of effective leaders

Analyzing Teaching and Learning Through Program Reviews

Kentucky's program reviews are perhaps the most unique part of its accountability system. The state has a history of examining the performance of students in multiple subjects using nontraditional assessment methods. As part of the

landmark 1990 Kentucky Education Reform Act, students participated in performance-based assessments that evaluated their achievement in a wide array of subjects. Those assessments were eventually replaced, but the program reviews in the Unbridled Learning accountability model are the state's current response to educator feedback that traditional standardized assessments are not always the most appropriate way to determine student learning in key areas.

The reviews are a multistep process. First, schools conduct their own reviews of their arts and humanities, practical living and career studies, writing, early learning, and world language programs. As part of this year-round process, they establish review committees for each program area that include teachers who work in the discipline, teachers from other content areas, and school leaders. The state also recommends that the review committees include parent and student representatives, other school staff, and relevant community stakeholders. The teams collect, analyze, and score evidence related to four categories: curriculum and instruction; formative and summative assessment; professional development and support services; and leadership support and monitoring. The examined evidence can range from curriculum resources and parent survey results to student performance data and examples of student work. Rubrics guide the scoring in each category and the category scores are totaled, resulting in a single number ranging between 0 and 12 for each program review. Programs scoring below 8 are designated "needs improvement," programs scoring between 8 and 10.7 are "proficient," and programs scoring 10.8 or higher are "distinguished." The review teams submit their findings to the school-based decision-making council, which then makes policy decisions based on the findings and submits the school reviews to the districts. Kentucky officials suggest that the review committees also share their reports with the entire school faculty. Once districts receive the school reviews, they evaluate them as part of their district improvement planning process and submit them to the state.

Kentucky has developed a proposed auditing process for analyzing and validating the program reviews. This spring, eight schools piloted the process, and the state is analyzing the pilot results to inform a final auditing design that will be recommended for statewide use. Based on the pilot results, it appears the schools are scoring their programs a bit too generously. Kentucky education officials are grappling with how to address this problem and one option they are considering is to provide districts and schools with clear examples of distinguished programs. They're also wrestling with whether and how to alter schools' program review scores based on audit results.

In the meantime, the state has provided numerous opportunities—online and in person—for educators to learn about both the overall program review process and the specific program review indicators in order to ensure as much consistency in their interpretation as possible. State officials emphasize that they don't expect schools to collect "crates and binders" of evidence to justify their scores and that they want the program reviews to become an ongoing mechanism for feedback and improvement for schools and districts.

Although the Unbridled Learning accountability system was first phased in four years ago, Kentucky officials have continuously worked to improve and refine the system. Right now, much state discussion is focused on the professional component since it will become part of accountability scores for the first time this upcoming school year. The exact measures to use in this area are still under consideration, and some are advocating for the educator effectiveness determinations to be enhanced with other measures, such as information about working conditions and professional development. The state is also closely watching the competency-based education and performance assessment work that is happening in New Hampshire and is considering the possibility of adding a performance-based component as a way for students to demonstrate college and career readiness.

"Ultimately, it comes down to two questions. What do you value? And how do you create a system that reflects what you value?"

Kentucky's accountability system includes a lot of measures, and all of those measures roll up into a single accountability determination for schools. This has led to a significant challenge—clearly communicating the results to parents and communities and helping them understand what goes into a school's final score. To help, the state has produced detailed school report cards, available online, where the public can see a school's results on every indicator, including the program review scores and college- and career-readiness rates. Sims concedes that it all amounts to a significant amount of information for schools and communities to digest but believes it's a worthwhile drawback to creating a more balanced system.

"Ultimately," says Sims, "it comes down to two questions. What do you value? And how do you create a system that reflects what you value?" In her view, the

biggest success of Kentucky's accountability model is that it values all subjects and recognizes how crucial they are to developing well-rounded citizens who can be successful in their next stages of life.

ALBERTA, CANADA

Building Capacity for Continuous Improvement

In September 2004, the Ministry of Education in Alberta, Canada, introduced its Renewed Funding Framework (RFF) for school authorities (the Canadian version of local school districts). The RFF is based on three pillars: funding—the equitable distribution of funds; flexibility—increased autonomy in the use of funds to facilitate local decision making to meet students' educational needs; and accountability—public accountability for use of resources and results achieved. When reflecting on the introduction of the RFF, Keith Bowen, director of system assurance for Alberta Education, recalls that most of the attention and conversation focused on the funding allocations and formulas, while the accountability pillar amounted to just one slide in the initial presentation of the framework. But that somewhat unremarkable beginning has resulted in potentially one of the most stable and long-running examples of multimetric school accountability currently in place.

System Profile
Minister of Education: David Eggen
https://education.alberta.ca/
Number of School Authorities: 378
Number of Schools: 2,374
Student Enrollment: 657,811

The accountability pillar was created to give Alberta's school authorities a consistent way to measure their success and gauge their progress using a broad spectrum of 16 measures. Together, these measures help Alberta's schools identify areas of strength and improvement and inform their priorities and strategies for continuous improvement.

The 16 measures are organized into the following seven categories: safe and caring schools; student learning opportunities; student learning achievement (grades K–9); student learning achievement (grades 10–12); preparation for lifelong learning, world of work, and citizenship; parental involvement; and continuous improvement (see Table 3). However, the province's focus is on the 16 measures themselves. Those measures include survey results; student learning indicator results such as dropout rates, high school completion rates, and the percentage of students eligible to receive Rutherford Scholarships (an indication of postsecondary readiness); and student academic achievement results on provincial achievement tests and diploma examinations.

Table 3: Alberta's Accountability Pillar	
Category	**Measures**
Safe and caring schools	• Teacher, parent, and student agreement that students are safe at school, learn the importance of caring for others, learn respect for others, and are treated fairly at school
Student learning opportunities	• Annual dropout rate of students aged 14 to 18 • High school completion rate • Teacher, parent, and student satisfaction with the opportunity for students to receive a broad program of studies, including fine arts, career, technology, and health and physical education. • Teacher, parent, and student satisfaction with the overall quality of basic education
Student learning achievement (grades K–9)	• Provincial achievement tests, acceptable standard • Provincial achievement tests, standard of excellence
Student learning achievement (grades 10–12)	• Diploma exam results, acceptable standard • Diploma exam results, standard of excellence • Rutherford scholarship eligibility • Diploma exam participation rate
Preparation for lifelong learning, world of work, and citizenship	• Teacher, parent, and student agreement that students are safe at school, learn the importance of caring for others, learn respect for others, and are treated fairly at school
Parental involvement	• Annual dropout rate of students aged 14 to 18 • High school completion rate • Teacher, parent, and student satisfaction with the opportunity for students to receive a broad program of studies, including fine arts, career, technology, and health and physical education. • Teacher, parent, and student satisfaction with the overall quality of basic education
Continuous improvement	• Provincial achievement tests, acceptable standard • Provincial achievement tests, standard of excellence

Many of these measures had been around prior to RFF's implementation, but the framework represented an opportunity to track them more consistently, strategically, and transparently. For example, the province launched its

Accountability Pillar surveys of Alberta students, parents, and teachers in 2004. Over 300,000 survey respondents now share their perceptions of the safety of schools, whether students are learning to care for and respect others, student access to a broad program of studies, the preparedness of students for work and citizenship, parental involvement, schools' commitment to continuous improvement, and more. The survey questions constitute several of the 16 indicators used to determine the performance of Alberta's school authorities.

In addition, the province instituted evaluations of each of the accountability pillar's measures, complete with performance labels and color coding. The evaluations are based on both a school authority's current results compared to the province's fixed standard for each measure as well as the authority's improvement on each measure over time. Although the evaluations are provided at both the individual measure and category levels, they don't roll up into a single score for a school authority. Bowen said this was very intentional. "Having a single score can serve an elemental desire from a communications and ranking perspective, but it accomplishes little in terms of system improvement. It's far more productive and actionable to show separate dimensions of performance."

It's worth noting that even without a single summative score, the accountability pillar's evaluative component received the most attention in the initial years of the framework because it introduced an element of objective analysis that school authorities were simply not accustomed to. But Bowen said that superintendents were generally more focused on putting the information into local context than on raising concerns that school authority performance was somehow being misrepresented by the methodology. And as time has passed and education leaders have become accustomed to the process, he says that they are primarily focused on sensibly interpreting and addressing their results.

From Implementation to Continuous Improvement

More than 10 years into the framework, Alberta has progressed from simply implementing its multiple measure accountability system to helping its districts and schools use the information for continuous improvement. Each school authority has the ability to slice and dice its data by any number of student groups and scenarios. "We're doing what we can to allow schools to focus on the "so what" part of the data, instead of draining their capacity on analysis," says Bowen. And putting data at the center of discussions between the Ministry's field services staff and school authority personnel has proven to be invaluable in examining priorities and identifying next steps. The accountability pillar has also

facilitated the identification of schools that have excelled in certain areas and collaboration among school authorities about what's worked well. Some school authorities have even built their own complementary data dashboards that add layers of nuanced and rich local metrics to the province's 16 common measures.

Alberta's Ministry of Education has refined the accountability pillar since its inception. While it has remained committed to the ultimate outcomes it desires for all students, advances in data collection and technology have allowed it to improve and tweak its measures as well as its survey questions. An area of ongoing focus is the engagement of all stakeholders—including students, parents, and community members—in schools' planning and improvement processes. "Our aim is to build school authority capacity for continuous improvement and ensure communities have a voice in the priorities that are being established locally," says Bowen.

"Our aim is to build school authority capacity for continuous improvement and ensure communities have a voice in the priorities that are being established locally."

CONCLUSION

All five of these featured examples vividly demonstrate the wide array of measures that can be reliably and fairly used in an accountability system (see Table 4). More important, the examples reveal that by using these multiple measures to evaluate district and school performance, the systems are sending a clear signal to their educators, families, and broader communities that certain aspects of student learning and well-being are essential and must be attended to. For example, New Hampshire's PACE pilot and Kentucky's Unbridled Learning accountability model are prioritizing a deeper and more nuanced look at student academic progress and understanding. New Hampshire is doing this by utilizing performance-based assessments that provide a more sophisticated and in-depth look at student learning, and Kentucky is accomplishing this through its program reviews that thoroughly capture the quality and breadth of schools' curricular and instructional offerings in several subjects. Tacoma, the CORE districts, and

Alberta, Canada, are prioritizing school culture and safety factors related to student engagement, well-being, and achievement through their use of school climate surveys.

Regardless of the exact mix of measures, each of the examples demonstrate how the inclusion of indicators beyond standardized tests supports a whole child approach to education, enhances and deepens understanding of student progress and well-being, and ensures that no single indicator of student achievement becomes synonymous with success (which would promote a singular focus on one aspect of learning and development to the exclusion of others). Although the featured examples have taken unique approaches to using multiple measures in accountability, their collective experiences provide five key lessons for those who want to adopt this more meaningful approach.

Start with the Child, Not with the Measure

Perhaps the most important lesson learned from these accountability systems is that—as important as they are—accountability measures should never be the starting point. Instead, leaders of these systems agree that the foundation of any accountability model should be the students themselves and what's needed to prepare them for their futures.

Alberta Education's Keith Bowen says that this is his primary takeaway for anyone interested in building a next-generation accountability system. "If you begin with a focus on measures, then you're missing the whole point. Start off by identifying what is important for your education system to accomplish over time and the outcomes you want for your students. Then use a backwards design process to identify measures, strategies, and the necessary resources to reach your goals."

Commit to Continuous Improvement

The value of multiple measures is using the rich and nuanced information they provide to drive ongoing improvement on behalf of schools and students. To maximize the effectiveness of multimetric accountability, communities must first ensure that the measures they are using align with local priorities, values, and goals. Then, they must continuously rely on the measures to inform their decision-making processes.

Tacoma's Joshua Garcia emphasizes that the measures should inform everything from policy actions and partnerships to decisions about resources and supports, rather than just using them as a mechanism for punishing low-performing schools.

Table 4: Accountability Metric Examples					
	Tacoma Public Schools	CORE Districts	New Hampshire PACE Pilot	Kentucky	Alberta, Canada
Standardized Test Scores	•	•	•	•	•
Graduation Rate	•	•	•	•	•
Preschool Enrollment Rate	•				
High School Readiness Rate		•			
Performance Assessment Results			•		
Social-Emotional Learning (SEL) Skills	•	•			•
School Safety/ Culture/Climate	•	•			•
Subject-Specific Program Offerings/ Quality				•	•
Family/Community Involvement	•				•
College/Career Readiness	•			•	•

Note: This table highlights the wide variety of measures being used by the five profiled systems in their accountability models, however, it is not a complete representation of all of the measures that they are incorporating.

The exact metrics and methodologies used within a single category can vary considerably. For example, Tacoma Public Schools includes each elementary school's annual summary of its Social Emotional Learning Action Plan in its accountability model, while the CORE districts rely on student self-reports of SEL skills and Alberta incorporates survey responses related to students' SEL skills. Similarly, metrics used in the school safety/culture/climate category range from tracking disciplinary incidents to analyzing school climate survey data. Alberta's family and community involvement measure examines teacher and parent satisfaction with parental involvement in decisions about their child's education. Meanwhile, Tacoma tracks percentages of registered volunteers in schools and percentages of expanded learning opportunities created with partner organizations, among other measures.

Although leaders of all of these systems are likely to agree that their accountability approaches are evaluating schools based on how well they're preparing students for life after high school, the final column of the table only denotes those systems that use specific and discrete college- and career-readiness indicators. These indicators include Tacoma's use of verified acceptance letters, Alberta's examination of Rutherford Scholarship eligibility, and Kentucky's college- and career-readiness rate.

Additionally, schools must be prepared to interpret and respond to their own performance on the measures, which means they need to be provided with the necessary expertise, capacity, and technology to analyze their results as well as the flexibility and support to make improvements.

Communication Is Key

"The work is sophisticated and comprehensive," says Tacoma's Garcia about the district's whole child accountability system. "Communicating it in a simplistic manner is an ongoing challenge for us. Our community continues to tell and teach us what they need."

New Hampshire's Leather agrees: "Whenever you're on the edge and in innovation land, it's not easy. A lot of education is involved, and you need to have public conversations in spades."

This need for communication and public engagement applies at all stages of accountability, from when the system is being developed to its implementation and refinement. Communicating the purpose and value of multimetric accountability and soliciting feedback about what measures to prioritize can promote understanding, engagement, and buy-in from the start. And thoughtfully communicating accountability results, including helping the public understand the importance and meaning of measures beyond more self-explanatory indicators like test scores and graduation rates, is especially important. Just providing parents and the broader public with an accountability score or even a set of data is insufficient.

Compromise Is Inevitable

All of the featured systems have had to make tough choices about what measures to include based on availability, capacity, and cost considerations. CORE's student self-reports of SEL skills are the perfect example. Although CORE district officials may have preferred to use teacher reports of such skills or performance-based assessments, the lack of capacity, the significant cost implications in terms of teacher time, and the absence of assessments focused on SEL skills limited their options. But instead of allowing those limitations to prevent such an important aspect of student growth and development to be reflected in its accountability system, CORE valued the information about students' SEL skills enough to develop a valid and reliable solution to at least meet the need until more sophisticated measures become available.

In short, the perfect should not be the enemy of the good when it comes to multimetric accountability. Systems should be able to take some calculated risks and experiment with new measures that they deem important, so long as they commit to ongoing analysis and refinement of their measures and are fair minded and sensible in their interpretation of results and how those results are used to evaluate schools and inform improvement.

There's No Finish Line

Building and implementing these systems is an ongoing process. Every one of the five examples, regardless of how far along they are in their work, is immersed in a process of analysis, refinement, and continuous improvement. "I've been at Kentucky's Department of Education for twenty-two years," shares Sims. "There hasn't been any point during which we haven't been involved in some sort of movement and evolution."

And despite this continuous improvement, no one claims to have built the perfect system. But leaders in all of the example systems agree that the challenges of using multiple measures in accountability are worth the rewards.

"If anybody thinks this type of accountability is less work, then I've got news for you," says Leather. "But if you're going to do this work, it may as well be the right work."

REFERENCES

Fullan, M. (2011). Choosing the wrong drivers for whole system reform. Centre for Strategic Education. Retrieved from http://edsource.org/wp-content/uploads/Fullan-Wrong-Drivers1.pdf

Conley, D. (n.d.). Four keys to college and career readiness. Retrieved from http://edimagine.com/four-keys/

Michelman, B. (2015). A lexicon for educating the whole child (and preparing the whole adult). *Policy Priorities, 21*(2).

Authors' Note

All data and information in this paper come from the websites and resources of the profiled states, districts, and provinces and from personal communication with each. Special thanks to Joshua Garcia, deputy superintendent, Tacoma Public Schools; Rick Miller, executive director, CORE Districts; Noah Bookman, chief accountability officer, CORE Districts; Paul Leather, deputy commissioner of education, New Hampshire Department of Education; Rhonda Sims, associate commissioner, Kentucky Department of Education's Office of Assessment and Accountability; Karen Kidwell, director, Kentucky Department of Education's Division of Program Standards; and Keith Bowen, director of system assurance, Alberta Education.