Introduction

In the last decade-plus, statewide accountability systems have emerged as a strategy for improving child outcomes, particularly for low-income and minority children. It is clear that state accountability systems have changed the behavior of schools. But to date, accountability efforts in both early learning and K–12 public school systems have not set the right goals for educators. Moreover, both early learning and K-12 have struggled to generate the capacity needed to improve schools at scale, and the strategies currently being used for improvement have frequently not had the intended effect.

For a state accountability system to truly succeed, it must both set the right goals and provide the right supports for achieving those goals. This paper proposes a policy framework that redefines statewide accountability based on the best elements of existing early learning and K-12 accountability systems in order to strengthen state approaches to educational improvement.

* In this paper, the term “accountability” draws from Martha Thurlow’s definition: “Educational accountability targets either the processes or results of education. A desired goal is identified and measures are identified for determining whether the goal is met.” Thurlow, Martha. (2009). “Accountability.” The Gale Group. http://www.education.com/reference/article/accountability/.

† In this paper, the term “school” refers to both K-12 public schools and school- or center-based early learning providers with an educational focus.
The broad goals for state accountability systems are widely agreed upon: Accountability systems are supposed to measure the professional practice of schools and then help schools improve their practices as a means of achieving better student outcomes. But to date no consensus has emerged about how to measure practice, how to help schools improve, or what student outcomes should be measured. While there is not yet clear national agreement on how accountability systems should work, there are promising trends to improve state accountability systems in both early learning and K–12 (which to date have been entirely separate systems). States can build on the best ideas in both early learning and K–12 accountability systems to create a single state education-accountability system from birth through high school—one that sets the right goals and identifies the supports needed to help achieve them.

In both early learning and K–12, recent federal efforts have helped catalyze state-level progress in goal-setting and the development of better supports. In early learning accountability, there has been an increasing focus on child outcomes as an essential goal, and greater attention to the need for metrics based on professional practice—metrics that can be linked to child outcomes. In K–12 accountability, there have been meaningful changes in what child outcomes are being sought. As for supports, the federal Early Learning Challenge grant pushed states to strengthen their early learning quality-improvement processes; at the K–12 level, federal waivers of the Elementary and Secondary Education Act (ESEA) have prompted shifts in thinking about school improvement. There is promise in these efforts, but they also have significant limitations.

While existing separate initiatives have led to incremental progress, states may be able to make greater advances by combining the best ideas of early learning and K–12.* Through this combination, states could create birth-to-high-school accountability systems that really do set the right goals and provide the right supports. The framework for a combined system could include states using the following elements in their accountability systems:

- **Performance metrics** that set the right goals for educators at both the early learning and K–12 levels. These goals will include metrics in two major categories: professional practice and child outcomes. Using professional-practice metrics will provide diagnostic information that can be used to improve practice, and child outcomes metrics can take a broader view of success than just proficiency in English language arts and math.

- **Performance measurements**† to track the metrics that include observations, which will provide information that school leaders and teachers can use to improve practice. This also will require child

---

* In this paper, the term “early learning” means school-based or center-based programs for children from birth through kindergarten entry that provide standards-based education; while many other programs for young children can and do provide early learning, this paper focuses on the subset of early learning programs that are or could be appropriately subjected to education accountability.

† In this paper, the term “metrics” means data that captures an outcome, and “measurements” means the method by which that metric is collected. For example, “student math achievement” is a metric, and “standardized math assessments” is a measurement by which that metric could be established.
outcome measurements that go beyond standardized assessment results to measure skills like executive function and cognitive processes that underlie lifelong learning.

- **Performance tiers** that provide transparent and easily understood information for parents about school quality but that are not used to determine punitive consequences for schools.

- **Supports** that are based on a detailed diagnosis of school performance and child need and are designed to improve professional practice and child outcomes. At the early learning level this may require fundamental shifts in resource allocation, given the number of children who currently do not have access to high-quality early learning.

Accountability systems are a promising approach because they have a demonstrated record of influencing decision-making at both the policy and practice level. If better accountability systems lead to improved decision-making about the use of resources, then they can play a key role in improving child outcomes. If a major purpose of early learning and K–12 accountability systems is to guide resource use in a way that improves educator practice and child outcomes, perhaps this moment of rapid change in both systems is an opportunity to bring the best thinking of each together to develop a unified accountability system that gives more complete information about what’s happening in our schools and helps drive the change needed to make things better.

This paper first lays out its proposed framework for how education accountability and support systems could work. It then discusses existing accountability and support systems in early learning and K–12, identifying specific ways existing systems need to be improved. It concludes by identifying steps to begin improving and unifying state education accountability and support systems.

---

* In this paper, the term “assessment” reflects the US Department of Education’s definition of “assessment” in the Race to the Top program: An assessment is “designed to measure a student’s knowledge, understanding of, and ability to apply critical concepts through the use of a variety of item types and formats.” US Department of Education. (December 2009). “Race to the Top Program Executive Summary.” http://www2.ed.gov/programs/racetothetop/executive-summary.pdf.
I) A Proposed Theory of Action for Birth-Through-High-School Education Accountability

State education-accountability systems can support improved outcomes, and it is clear that they have the power to significantly change behavior. Because state-level metrics drive local-level action, states need a set of metrics that provide local actors with the right incentives. That will require using new metrics and measurements that capture a better set of desired outcomes than the outcomes currently driving state accountability systems. States also need to report results in ways that better inform the public and act on results in a tailored manner that leads to improvement—focusing state resources on helping the children most likely to need dramatic change in their schooling to succeed. An accountability system of this kind would not only change behavior, it would lead to improved outcomes for students. This paper is meant to offer a starting point for discussion of what a more effective accountability system might look like.

A) The Right Performance Metrics

Because accountability metrics drive behavior, it is essential that the metrics be focused on the right goals. Stated broadly, the goal of the education system is that children achieve positive outcomes and that they are supported in doing so by adults engaged in high-quality professional practices. This high-level statement addresses both the end (child outcomes) and the means (professional practices). For states to develop metrics that support this statement of the goal will require a combination of metrics that address the ends and the means, not just one or the other.

Existing accountability systems have suffered from a tendency to focus on only outcomes or practice and to take an inappropriately narrow approach within their area of focus.

- K–12 accountability has focused on child outcomes but has defined those outcomes very narrowly. It relies primarily on outcomes measured by standardized tests in a limited range of subjects—which is an indirect way to improve the practice of educators, and has also meant that strong accountability results are frequently highly correlated with wealth. Given the limited scope of standardized testing for accountability purposes, there is also significant concern that the accountability system focused solely on child outcomes is not even fully capturing the child outcomes desired from the system.

- Early learning historically has focused on professional practices through input metrics—that is, metrics that do not directly measure child outcomes (although some are linked to improved child outcomes through research). But while early learning accountability has focused on adult practices, in many instances it has done so through checklists of activities that are not necessarily linked to improved teaching. This has resulted in programs improving their accountability ratings without either improving their teaching practice or producing better child outcomes.

New accountability systems should include metrics addressing both child outcomes and professional practice—and in doing so can build on lessons learned from ongoing improvement efforts in both early learning and K–12.
The Five Essential Supports for School Improvement (referred to as the “five essentials” or “5E”) are based on longitudinal research by the University of Chicago Consortium on Chicago School Research that determined five key ingredients or “essential supports” to school success. Schools strong in most of the essential supports were at least ten times more likely to achieve significant academic gains and experience improved student attendance and engagement compared with schools showing strengths in only one or two areas. The 5E are:

1. **A coherent instructional guidance system**: Schools have strong classroom curriculum alignment where instruction is academically challenging, engaging students, and provides adequate tools for teachers to advance learning.

2. **Professional capacity**: Schools recruit quality teachers and staff and provide ongoing professional development that fosters strong commitment to improvement and capacity to work together.

3. **Strong parent-community-school ties**: Schools create a welcoming environment for parents and families, and build partnerships with local community institutions.

4. **Student-centered learning climate**: Schools provide a safe, stimulating, and nurturing learning environment for all students.

5. **Leadership to drive change**: The principal actively works with teachers to develop a clear, strategic vision for school success that is inclusive and focused on improving instructional practice.

The five essentials provide a research-based framework for whole-school improvement. The 5E assessment tool, which draws from the research framework, reliably measures and predicts which schools will demonstrate success or stagnate over time. The state of Illinois is now using the 5E as part of a statewide survey of students and teachers. State legislation in 2011 required the Illinois State Board of Education (ISBE) to implement a survey that captured the learning conditions and climate of schools. ISBE partnered with the University of Chicago to develop the Illinois 5Essentials Student Survey for students in grades 6–12 and the 5Essentials Teacher Survey for pre-kindergarten–12th grade certified teachers. While school districts in other states have used the survey, Illinois is the first to administer it statewide across all districts. The survey results are intended to help frame and direct school improvement planning.

---

1) **The Right Performance Metrics: Professional Practice**

If a major goal of accountability is to lead to improved practices in schools, then accountability systems should actually measure school and professional practice directly. There is an emerging literature about what school practices are most important for success. A landmark study by the University of Chicago Consortium on Chicago School Research identified five essential supports for school improvement (referred to as the “five essentials” or “5E”)—which are organizational features of schools that, when present, improve the chance of positive student outcomes. These features are:

1. A coherent instructional guidance system
2. Professional capacity
3. Strong parent-community-school ties
4. Student-centered learning climate
5. Leadership to drive change

Research has shown that a substantial, sustained weakness in even one of the essential features significantly threatens the ability of the school to improve—and if three or more of the five essentials are strongly present, the school is at least ten times more likely to improve over time. These supports can provide a starting point for states in developing research-based metrics of professional practice. Research by the Ounce of Prevention Fund and the Consortium on Chicago School Research is also under way to specifically apply the 5E in the early learning context.
States may also want to work on aligning their school accountability metrics with the practice metrics used in their statewide teacher evaluation systems, given the extensive work states have done in recent years to improve teacher evaluation. Interestingly, at the K–12 level, the work on teacher evaluation has been in some ways the reverse of what has been proposed here—in teacher evaluation, states have sought to integrate child-outcome metrics to a process that historically has not included them, as opposed to adding professional-practice metrics to a process that has been based solely on child-outcome metrics.

States will undoubtedly need to proceed carefully in this area, given that the use of metrics focused on professional practice has not been a focus of accountability systems in the past. In doing so, states should be mindful that there is a value to minimizing the number of professional-practice metrics: If there are too many metrics for schools to easily keep track of, all of the metrics will lose value. This has been a key lesson in some early learning accountability systems where the sheer volume of requirements minimizes the importance of any individual requirement, even when some requirements are clearly more important than others.

2) The Right Performance Metrics: Child Outcomes

Child outcomes should be an important part of any state accountability system, but “child outcomes” and “child standardized test scores” are not the same thing. For example, at the high school level, graduation rates are a form of child outcome that is already included in accountability systems. The positive outcomes that policy leaders and parents want from the education system are broader than just high scores on standardized tests, and the child outcomes for which we have accountability metrics should be broader than standardized tests to reflect that. A move to a new accountability system could force a productive conversation about exactly what child outcomes states desire from and prioritize for their schools; clearly, English-language arts and math achievement are important, but just as clearly they are not the only outcomes a state might want from its education system. States also might reasonably differ between themselves on what outcomes they want, with some states emphasizing English-language arts and math and others focusing on a broader range of metrics.

The right child outcomes metrics will likely vary across age spans. For example, even if there were no federal requirement to do so, many states would likely include high school graduation rates in their high school accountability, but it is likely that no states would include high school graduation rates in their elementary school accountability. In the earlier years, child outcomes should be tied to developmental milestones, with recognition that in the younger years children develop different skills at different rates and that wide variance is appropriate and expected. In that spirit, it will be important for child outcomes to incorporate child growth, which is likely to be a better measure of school performance than proficiency. Overall, the approach to child outcomes as accountability metrics should continue evolving toward a broader range of outcomes that focus increasingly on growth and go beyond assessment scores.
Unfortunately at this time, neither early learning nor K–12 offer significant lessons in how to mix a range of metrics reflecting different desired child outcomes. States will need to identify the best possible child-outcome metrics and balance those with professional-practice metrics. States may choose to have different balances for different age spans—most likely with a gradual shift in balance toward child outcomes as children get older.

**B) The Right Performance Measurements**

If an accountability system is going to include metrics focused on professional practice and child outcomes, then the measurements it uses must address both. Measures that are meant to capture whether schools are engaged in high-quality professional practice require external observation—which to date has not played a significant role in K–12 accountability, but has played a meaningful and growing role in early learning accountability. As for child outcomes, historically, standardized tests have been the dominant child-outcome measurements, despite widespread discomfort with their role. In measuring both professional practice and child outcomes, there are lessons to be learned from existing practice here and abroad that can help update measurements to use in state education-accountability systems.

**1) The Right Performance Measurements: Professional Practice**

Measuring professional practice for school accountability is already common practice in England, where a school inspectorate is a key part of school improvement. Since the early 1990s, the English public school system has put in place independent inspectors to evaluate the overall quality of its schools in an effort to hold them accountable and promote improvement while also offering parents more information. External inspectors make on-site visits generally once in each three- to six-year inspection cycle, although the frequency of inspection varies based on past school performance. The inspections typically occur on short notice so as to limit last-minute school preparations for the visits. Inspections focus on reporting student learning progress, the quality of instruction, the behavior of students and safety climate of the school, and the effectiveness of school leadership and management. The inspections utilize a blend of metrics to assess school performance, including test scores, classroom observation, and staff, student, and parent feedback. Importantly, these areas of focus are aligned with those described in the five essentials. To measure learning and development in early learning, inspectors are also advised to study a best practices framework for child development that lays out critical areas of learning for young children (noting that these are not a prescribed set of expectations).

The inspector decides overall school performance based on a grading scale of (1) outstanding, (2) good, (3) requires improvement, and (4) inadequate; schools receiving a failure rating may be subject to increased and intensive inspections. After inspections are conducted, inspectors provide schools feedback about areas that need improvement at the student and school levels and offer suggestions on what can be done to improve. Information is then released to the public on the Internet through a full report that includes a description of the school, the inspection’s findings, and the school’s overall grade. The results from these inspections
have been shown to have overall positive impact on school improvement with regard to higher academic achievement, quality of instruction and supports, improved management and leadership, and enhanced connections among teachers and students.\textsuperscript{28}

While the English system may be the best known in the United States, inspection-based accountability is also used elsewhere. For example, the Dutch education system has implemented an inspectorate process for assessing schools in a way that shares information with the public about school performance and identifies poor performing schools in need of additional inspections. The goal is not to penalize the school but to respond to the inspectorate findings by providing resources to build school capacity to overcome the barriers identified.\textsuperscript{29}

While English- and Dutch-style inspections are not currently a feature of K–12 accountability in the United States, the practice of external review is already common in US early learning accountability:

- As part of child care licensing, reviewers typically observe early childhood programs for compliance with basic structural and environmental quality features including health, safety, and physical environment.\textsuperscript{30}

- Accreditation of child care programs—which is voluntary and aims to meet higher standards than licensing rules—generally includes a site visit by an external observer in addition to a self-assessment component.\textsuperscript{31}

- Trained independent assessors have also been used by state quality rating systems to make on-site visits and administer standardized assessment tools like the Classroom Assessment Scoring System (CLASS) and Environment Rating Scales (ERS) to rate and verify a program’s level of quality (see sidebar, next page).

- States have used external observers to measure preschool quality and support improvement efforts.\textsuperscript{32}

- Head Start programs are also subject to monitoring reviews that require assessors to use research-based observations for measuring dimensions of program quality, such as classroom environment, teacher-child interaction, and teacher instructional practice.\textsuperscript{33}

Over time, some states have moved from assessors reviewing a checklist of basic aspects of quality to evaluating key elements of quality care and instructional practice.\textsuperscript{34}

An important assumption underlying this approach is that quality teaching can be defined and measured.\textsuperscript{35} This is an area where state efforts to develop new teacher evaluation systems may be instructive, and there are teacher evaluation rubrics that can be used and improved upon.\textsuperscript{36} (See sidebar, page 10) In measuring the quality of teaching, it is important to note that instructional practice must be age-appropriate, and that in early learning, instructional practice includes substantial play-based teaching and learning.\textsuperscript{37}

Accreditation processes are another possible model for the measurement of professional practice...
quality in schools serving all age ranges. Even though external review is not generally required at the K–12 level by state accountability systems, many schools use outside accreditors to evaluate their professional practice. In early learning, quality rating and improvement systems (QRIS) can give credit to programs for being externally accredited, in some instances by automatically assigning them high program ratings. While K–12 schools may not receive an equivalent benefit, some of them still use accreditation as part of a continuous quality-improvement process.

An accountability system based on outside observations of schools and educator practice would be more expensive than existing accountability; maintaining a staff of well-trained observers is an expense states currently are not bearing. States can mitigate this to some degree by focusing observations on the schools likely to need them most, but this work cannot be done without repurposing resources. If improved accountability leads to improved graduation rates and college success, then the up-front costs will have a long-term payoff. But it must be acknowledged that in many states the up-front cost of measuring performance would be a barrier.

2) The Right Performance Measurements: Child Outcomes

Child outcomes as measured by standardized tests are the key drivers of accountability in K–12 education, with high school graduation rates also playing a role in high school accountability. These metrics have been used in part because measurements for them are readily available; indeed, in some instances it appears that metrics are used primarily because they are easily measured, not because they meaningfully capture student progress or predict later learning.

Classroom Assessment Scoring System (CLASS) and Environment Rating Scales (ERS)

The Classroom Assessment Scoring System was developed at the Curry School Center for Advanced Study of Teaching and Learning at the University of Virginia as part of a national study on early learning. The CLASS, currently distributed by Teachstone, is an observational tool used to assess classroom quality in grades pre-k to 12. It is currently implemented by the Office of Head Start and also in state and local efforts to measure and enhance early education and care programs. The instrument—which includes versions specifically for infants and toddlers as well as preschoolers—focuses primarily on measuring teacher behaviors and teacher-child interactions shown to be linked to student achievement and development. The CLASS categorizes teacher-student interactions into three domains (emotional support, classroom organization, instructional support) and also further organizes interactions into multiple dimensions. These are dimensions of teaching and learning with an empirical link to academic and social outcomes for children. The CLASS tool has been validated and tested reliable in more than 3,000 classrooms.

The Environment Rating Scales, designed by the Frank Porter Graham Child Development Institute at the University of North Carolina at Chapel Hill, assess the overall process quality of early learning and school-age care programs. The observational tool evaluates quality by measuring physical learning environment in a broad sense. The instrument, researched-tested as valid and reliable, focuses on assessing the following: physical environment, basic care, curriculum, interaction between staff and children and among children themselves, schedule and program structure, and parent and staff education. ERS is currently implemented for program improvement and evaluation purposes related to early learning licensing, accreditation, credentialing, and quality-rating systems.
If states have a conversation about which metrics matter to them, they may find that of the metrics that matter, most do not have measurements attached to them—for example, civic engagement or art and music proficiency. A new state of the art is emerging in accountability measurements, but it may take states time to develop measurements that meet their needs.

Early learning accountability systems have not historically included child outcomes, although leaders in the field have called for it to do so. Leading researchers in the field have written that it is inappropriate to use child-assessment results for program accountability with children 8 years old or younger. States seeking to include child outcomes in early learning accountability should therefore identify child-outcome metrics that can be measured through developmentally-appropriate measurement tools. In some cases, states may identify appropriate measurements that are already available that they would like to include in early learning accountability; in other instances, they may not, creating pressure to develop new measurements. This too is an area where further development is needed.

**Best Practice Pedagogy in Early Learning**

A large and growing body of research demonstrates that early learning educators’ approach to interacting with students in the classroom is highly predictive of children’s achievement. Children in the early childhood years learn within the context of trusting, supportive relationships—with caregivers, parents, and teachers. Research therefore shows that effective practice in early education involves positive and emotionally supportive teacher-child interactions in which teachers demonstrate respect, responsiveness, sensitivity, enjoyment, and regard for students’ perspectives.

Effective practice is also language- and content-rich and intentional, with a focus on developing young children’s higher-order reasoning skills through the use of evidence-based, developmentally appropriate curricula and strategies focused on interactions between teacher and child, including open-ended questions, conversation, modeling, variety, play-based engagement, and promotion of students’ interests. These interactions help facilitate joint attention—paying close attention to what has captured a child’s interest and staying with the child’s interest through commentary and shared physical manipulation of classroom tools. Joint attention not only supports language and vocabulary development but is also an important way adults show children they are interested in them and their ideas. Infants and young toddlers might not be able to say what they are interested in, so effective teachers in infant and toddler classrooms need to be keen observers, paying careful attention to what children look at, react to, and point to—the behaviors that communicate their interests, likes, and dislikes. Effective teachers follow a child’s eye gaze and talk specifically about what the child is looking at.

Building on this research, effective, comprehensive early learning curricula often structure time in the day for preschoolers to act on their interests by choosing areas in the classroom to play, materials to incorporate into their play, and scenarios to act out in their play with other children, under educators’ guidance. During these particular times in the day, teachers create episodes of shared attention with preschoolers, following their ideas, interests, and scenarios, and offering commentary, content-specific vocabulary, and questions to help children expand the complexity of their vocabulary, syntax, and play. Effective early learning educators gently guide children’s learning by populating the classroom with enriching materials and facilitating engagement with peers and adults, creating opportunities for exploration, discovery, and social development. While providing ample opportunities for such child-directed activities, early learning classrooms should be well-organized, with clear expectations, efficient behavior management, and predictable routines and transitions.
C) The Right Performance Tiers

State accountability systems at both the K–12 and early learning levels sort schools into performance tiers, which serves several primary purposes. One is the important value of providing information for parents and the public about the quality of programs. The standards for inclusion in higher tiers set expectations for schools by defining a pathway to improvement and what is supposed to be achieved. A somewhat different purpose has been that tiers organize the provision of incentives and supports for schools, with different incentives and supports tied to placement in particular tiers.

Changing the metrics of accountability does not substantially change how states should approach organizing tiers for the purpose of informing the public and setting expectations for schools. Incorporating professional-practice metrics should increase the connection between performance tiers and the actual quality of instruction being provided. However, setting performance tiers and weighing multiple metrics is already a complex exercise and is in many ways more art than science. It may be reasonably straightforward for states to identify the schools in their top performance tier as those with both outstanding professional practice and excellent child outcomes—and it may also be reasonably straightforward to identify the schools in their lowest performance tier as those where results are just the opposite. Sorting out schools in the middle is already a challenge, however, and even with improved metrics it will likely take some time for states to sort out how best to differentiate among schools in the middle tiers.

Importantly, the use of the professional-practice metrics in accountability should relieve pressure on performance tiers to serve their other function, identifying school support needs. Existing systems group together into tiers schools with similar test scores that may have very different challenges because existing accountability measures do not identify what practices a particular school needs to improve. When consequences are based on tiers, the consequences are not necessarily well suited to the problems a particular school faces—to date, in many instances, there has been a misalignment between the reasons a school was identified for interventions and the interventions actually being used. School inspections would provide much richer information about what behaviors need to be changed and would allow states to mobilize supports to address the actual problems. Consequences based on performance tier have been a relatively blunt instrument; integrating professional practice into accountability should substantially increase the precision with which supports are designed to lead to improved professional practice.

D) The Right Supports

Developing the right system of supports in states will require an ongoing process that integrates substantial feedback from the field, but a focus on practice-oriented metrics and measurement increases the chances that the right supports will be mobilized. Diagnostic information from school observations can be useful to improvement whether or not the school becomes the recipient of state improvement assistance. States will likely be most effective and efficient if they focus their resources
In Early Learning, Supports Are Necessary But Not Sufficient

Supports will be essential to education accountability in both early learning and K–12, but it is important to acknowledge that while children are guaranteed access to K–12 public schools that are designed to provide an educational experience, in early learning there is no such guarantee. Accordingly, many children do not have access to early learning at all. Moreover, historically many of the programs serving children birth through age five have not been designed or funded sufficiently to provide a high-quality educational experience.

In recent years, significant progress has been made in providing educational opportunities to young children. The need for high-quality early learning is well established. All states now have learning standards for 3- and 4-year-olds, and most have learning standards for infants and toddlers, which establishes a framework for birth-to-five education. However, the funding stream serving the most children in this age group—federal and state child care dollars—is primarily a work support and is neither designed nor funded to provide standards-based education. Even many state pre-k programs are not funded adequately to provide a quality educational experience. The low per-child funding rate for these programs has made it difficult to deliver high quality, in part because this funding approach remains insufficient to hire and retain the best teachers.

Because many early learning programs have not been provided adequate funding to support high-quality professional practice, the imposition of true educational accountability on early learning systems is likely to show that the quality of professional practice needs substantial improvement in many early learning programs. Given that, we strongly believe that states should do more than just provide targeted assistance to early learning programs that receive low ratings in accountability systems—they should fundamentally change how they fund early learning to ensure that programs provide a high level of professional practice, particularly programs serving children who without state support would have no access to standards-based early education. While that fundamental change in how states approach early learning is by no means an inevitable outcome of shifting to a birth-through-high-school education-accountability system, we are hopeful that it would be a major result of increased education accountability in the pre-kindergarten years.

* State pre-k programs are generally defined as state-funded, standards-based education programs serving 3- and 4-year-olds.

on the schools with the lowest professional capacity, which the new metrics will actually identify. States can also then focus on ensuring that any resources provided align with core issues of a school so that they meaningfully address the challenges at hand. Ideally, supports are based on the diagnosis of need at a school, are designed to build the capacity of the school to address those needs, and include a focus on improving learning and instruction.

While the focus of accountability should be on supports, those supports must be backstopped by consequences. Many schools given awards to undergo major improvement still face challenges to making significant gains. Both early learning and K-12 accountability systems have included different provisions for significant consequences when schools are incapable of improving even with support. Punitive measures should be used only after schools have been given a reasonable chance to improve and failed to do so. However, if a school cannot perform even after receiving assistance, then it is appropriate to insist on major changes in school management for the school to continue receiving government funding.
Integrating professional practice into accountability would improve the ability of states to identify and provide the right supports for schools, shifting the focus of accountability systems from identifying problems to solving them. But it must be acknowledged that problem solving at the state level is not an exercise that has historically been successful at scale. Indeed, state-level problem solving is often hampered by the fact that the term “state” oversimplifies in early learning what is often a complex web of agencies and offices; early learning governance is frequently spread across multiple agencies,\(^7\) and low-performing school districts are often subject to oversight by multiple offices of the state education agency,\(^6\) meaning that “state” assistance is not always so simple to mobilize. Moreover, state offices typically have insufficient capacity to help the full range of schools that need assistance, and frequently lack the capacity to develop the nuanced understanding of individual communities necessary to be an effective partner to schools in a wide range of local contexts. States also have not always designed supports that account for the fact that improvement efforts generally will not succeed without strong school-level leadership. As states seek to build their capacity to help local schools, they should focus on what is actually achievable from the state level given their role, and what state-level changes will be needed to accomplish that which is achievable.

### PROPOSED BIRTH TO HIGH SCHOOL EDUCATION ACCOUNTABILITY

<table>
<thead>
<tr>
<th>Ages 0–3</th>
<th>Ages 3–5</th>
<th>Grades K–2</th>
<th>Grades 3–8</th>
<th>High School</th>
</tr>
</thead>
</table>
| **Metrics** | • A mix of metrics balancing child outcomes with professional practice  
• Child outcomes will vary across age spans  
• Balance between child outcomes & professional practice will vary across age spans |  |  |  |
| **Measurements** | • Measurements of child outcomes that are research-based and age-appropriate  
• School observations used to measure professional practice |  |  |  |
| **Tiers** | • Tiers that communicate to the public the quality of outcomes and practice at a school  
• The highest tier will be reserved for schools that score highly on both, and the lowest tier for schools that score poorly on both  
• Over time it will become possible to draw more meaningful distinctions in the middle tiers |  |  |  |
| **Supports** | • Supports to schools designed to address issues identified by the observation of professional practice |  |  |  |
II) What’s Happening Today in Education Accountability

Having broadly defined what birth-through-high-school education accountability and supports ought to look like, it is important to consider how existing accountability systems match up with that definition. Early learning and K-12 accountability are currently two entirely separate systems, both of which are evolving rapidly. While they are evolving in ways that bring them marginally closer to the model defined above, both still have a long way to go—and in most states, they are evolving separately. Indeed, in many states, child care and the accompanying QRIS are administered by a state agency other than the state education agency responsible for K-12 accountability. This only reinforces the divide between these two systems.

This section gives an overview of key themes in early learning and K-12 accountability. Because these systems are administered separately in most states, there may be few leaders in positions of authority who are intimately familiar with both. This overview is intended to help leaders and advocates understand similarities and differences between the two systems and identify areas in which each could evolve to better reflect the principles articulated in section I.

**KEY STRUCTURAL DIFFERENCES FOR ACCOUNTABILITY BETWEEN EARLY LEARNING AND K–12**

<table>
<thead>
<tr>
<th>Early Learning</th>
<th>K–12</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metrics</strong></td>
<td></td>
</tr>
<tr>
<td>• Focused on inputs and structural criteria</td>
<td>• Limited range of student outcomes</td>
</tr>
<tr>
<td>• Not directly tied to child outcomes</td>
<td>• Not directly tied to practices, processes, or instruction</td>
</tr>
<tr>
<td><strong>Measurements</strong></td>
<td></td>
</tr>
<tr>
<td>• Objective criteria</td>
<td>• Assessments</td>
</tr>
<tr>
<td>• Observation measures</td>
<td>• Graduation rates</td>
</tr>
<tr>
<td><strong>Tiers</strong></td>
<td></td>
</tr>
<tr>
<td>• Point, block, or hybrid/combination</td>
<td>• Reward, priority, and focus schools</td>
</tr>
<tr>
<td>• Letter grades</td>
<td>• Letter grades</td>
</tr>
<tr>
<td><strong>Supports</strong></td>
<td></td>
</tr>
<tr>
<td>• Typically financial incentives based on child care reimbursement</td>
<td>• Focused on building capacity and changing practices</td>
</tr>
<tr>
<td>• Professional development and training</td>
<td>• May be a governance change</td>
</tr>
<tr>
<td>• On-site technical assistance</td>
<td>• May be focused on a specific provider (school) or on a more systemic level (district)</td>
</tr>
<tr>
<td>• Improvement grants</td>
<td></td>
</tr>
</tbody>
</table>
A) Early Learning Accountability

In many ways it is misleading to describe “early learning accountability” as if it were a single enterprise. Historically, Head Start, state pre-k, and child care programs have all been subject to different accountability. Head Start has a thorough set of program standards, and programs are monitored to determine whether they are meeting those standards; the lowest-rated programs are now required to recompete for their grants. State pre-k programs generally have design requirements of their own that funded programs are required to meet, but states have not historically provided public rankings of pre-k program quality. Child care accountability varies substantially from state to state, but in recent years an increasing number of states have used QRIS to grade and rank programs, with incentives built in for program improvement.

CURRENT ACCOUNTABILITY OVERSIGHT STRUCTURES FOR K-12 AND EARLY LEARNING

<table>
<thead>
<tr>
<th>K-12</th>
<th>Head Start</th>
<th>Child Care</th>
<th>State Pre-K</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEDERAL</td>
<td>Department of Education</td>
<td>Office of Head Start, Department of Health and Human Services</td>
<td>Office of Child Care, Department of Health and Human Services</td>
</tr>
<tr>
<td>STATE</td>
<td>SEA</td>
<td>State Human Services, SEA, or Early Learning Agency</td>
<td>SEA or State Human Services Agency</td>
</tr>
<tr>
<td>LOCAL</td>
<td>LEA</td>
<td>Local “super grantee”</td>
<td>District</td>
</tr>
<tr>
<td>SCHOOL</td>
<td>Public School</td>
<td>Public School or Private Provider</td>
<td>Private Provider</td>
</tr>
</tbody>
</table>

Note: The chart above outlines the general accountability oversight structures currently in place for K-12 and early learning. In early learning, individual providers often use multiple funding streams and are therefore held accountable to multiple agencies. Also in the chart above, a local “super grantee” refers to a grantee overseeing multiple Head Start delegate agencies that contract services. For example, in 2013, the Los Angeles County Office of Education contracted with 24 school districts and private nonprofit agencies operating more than 300 sites to provide early learning services including Head Start and Early Head Start. The dotted line indicates that in some state pre-k programs, accountability flows directly from the state to schools.

Quality rating and improvement systems are the early learning accountability methodology with a framework most like K-12’s, and not coincidentally are also the methodology growing rapidly across states and early learning programs. A significant driver of QRIS development in the last few years has been the federally-funded Race to the Top-Early Learning Challenge. The Challenge began in 2011 as a competition

---

* While the term “quality rating and improvement system” is used to describe an overall approach, most individual state systems are given specific brand names, such as Better Beginnings (Arizona), Paths to QUALITY (Indiana), Keystone STARS (Pennsylvania), and BrightStars (Rhode Island).
for states to develop early learning systems, with QRIS a major focus of the grant. As defined in the Challenge, QRIS use the framework of accountability described above: It has metrics and measurements meant to sort providers into performance tiers, with the tiers having consequences (both positive and negative) for performance. The competition rewarded states that committed to expanding QRIS to preschool and Head Start. In total, 20 states have been awarded over $1 billion dollars to conduct their work.

The changes to QRIS brought through the Early Learning Challenge have in some ways moved states closer to the right approach to accountability in early learning. But the metrics used in QRIS still have an insufficient focus on pedagogy, and the support infrastructure is still lacking. For early learning to reach its full potential to improve long-term outcomes, both of these issues must be addressed.

1) Early Learning Accountability: Current Performance Metrics

Current metrics used in quality rating and improvement systems generally focus on program quality components that are externally observable (which can include document reviews). In many cases, the lowest level of QRIS is tied to child care licensing requirements, which are typically focused on health and safety. Common criteria include staff qualifications and professional development, curriculum and learning activities, administration and business practices, family engagement, staff-child ratios and group size, the use of child assessments, and health and safety. “Teaching and learning” is often a component of the ratings—which typically include the quality of instruction, learning environment, and teacher-child interactions.

Many of these metrics are rooted in child care, but QRIS are evolving to draw criteria from multiple programs (child care, Head Start, and preschool) to facilitate alignment and increase participation. Head

The Scope of QRIS Coverage

In those states that have QRIS (and not all do) it is frequently a mark of some quality for schools to be included in QRIS even at the lowest tier, as some programs—even programs that receive state support—do not achieve that level of quality. Moreover, participation in QRIS is frequently voluntary. Because QRIS include programs under multiple funding streams (child care, Head Start, and state pre-k)—each of which may have different QRIS participation rates in a given state, to say nothing of across states—the scope of QRIS’ coverage is not at all comparable to the scope of accountability coverage in K-12.

Of course, some difference in the scope of coverage is to be expected given the different natures of the systems. Public K-12 systems are obligated to accept all children who want to attend them, and compulsory attendance laws require children to attend; in early learning, schools have no such obligation, and parents have no obligation to send their children to school. In addition, early learning takes place in a wide variety of settings, many of which draw on multiple funding streams (including parent payments).

States are currently working to expand the scope of their QRIS coverage even as they try to improve the structural elements of QRIS. This expansion work would likely need to continue if states moved from redesigning QRIS to creating a unified education-accountability system.
Start programs are able to participate in close to all statewide QRIS, and state-funded pre-kindergartens are eligible in approximately two-thirds of QRIS. In some states, pre-kindergarten programs are required to reach specific levels of quality within QRIS to receive public funding. States have also made special efforts to ensure that their QRIS reflect the developmental needs of children from birth through age three. States including infant and toddler quality indicators have integrated specific infant and toddler training and professional development, appropriate teaching and curricula for birth to age three, lower staff-child ratios and group size, increased family involvement and information sharing regarding child routines, program policies that promote continuity of care, and use of assessments targeted for infants and toddlers.

Metrics measuring the qualifications of personnel have played an increasingly essential role in QRIS, but challenges remain to building a sustainable, high-quality early learning teaching force. The education level of the early learning teaching force is far below that of K–12 teachers: In K–12, essentially all teachers have at least a bachelor’s degree, while in early learning, such degrees are not required in Head Start, some state pre-k programs, and child care. Supports such as professional development, mentoring, and paid training may be more readily available in K–12, and many community-based early learning programs lack the infrastructure and adequate funding to offer benefits that enable teachers to effectively apply their knowledge and continue to develop their practice. Additionally, differences in compensation and perceptions of the role of teachers in K–12 compared to early learning have contributed to persistent issues facing the early learning workforce, such as high turnover rates and limited career opportunities in the field. Changes in state accountability systems to focus on professional practice could help to drive improvement in these areas.

While in some states the existing QRIS are relatively complex, there have been efforts to move toward simplifying the metrics to focus on a handful of simple but powerful standards. In addition, the Early Learning Challenge required states to conduct validations of QRIS to determine whether they had the intended impact. These forces could combine to make the metrics of QRIS evolve relatively rapidly over the next few years.

2) Early Learning Accountability: Current Performance Measurements

One key difference in performance measurements between early learning and K–12 is the performance of children on summative assessments. There are no assessment-based output criteria suitable for use in the QRIS years. Accordingly, QRIS have included a more substantial focus on process.

As states and localities have developed and adopted QRIS, these systems have focused on improvement around early learning inputs and child outcomes. The general underlying theory is that higher quality programs will contribute to improved child functioning. Rather than using one main measurement

---

to assess program performance, QRIS typically incorporate multicomponent assessments designed
to measure process quality and make quality programs more transparent and better understood.
These assessments include capturing the level of teacher-child interaction through the use of the
CLASS observation tool110 and assessing quality of environment, materials and activities, and caregiver
interaction using ERS.111 To support a process of continuous improvement, QRIS provide programs
feedback and technical assistance based on assessment findings and offer a variety of incentives to
make enhancements.112

3) Early Learning Accountability: Current Performance Tiers
QRIS typically use star rating levels (one star through five star), but they also use other terminology to
describe their tiers.113 No state currently uses letter grades to describe its tiers.114 The goal of the rating
system is to make it easy for families to understand the quality of programs to help guide their choices
for enrolling their children.115

There are three primary methods for sorting programs into tiers: point systems, block systems, and
hybrid systems.116 These methods operate as follows:

• **Point systems:** A program earns points within each category of standards, with the total point score
determining the level of rating. For example, a program is awarded points in each quality criterion
and must reach a minimum number of total points to achieve the level and move to the next—but
can achieve that overall point total even if the program is substandard on some criteria. Additional
points may be provided in some states if a provider attains outside accreditation.117

• **Block systems:** A program must meet all standards in each quality level to progress to the next level.
For example, a program can move to level 3 once it achieves all of the requirements for levels 1 and
2.118 Many states require a program to meet licensing standards to be included in the system at level 1.

• **Hybrid or combination systems:** A program is rated using the block approach and point system.
The initial levels are blocks, while the higher levels are achieved through a point system. For example,
a program might have to meet all standards in level 2 to advance to level 3 but could advance to level
4 by acquiring enough points in some criteria without scoring highly in others.119

Different states use different systems, with the most common form of assigning ratings being the
block system. According to the National Center on Child Care Quality Improvement and the QRIS
National Learning Network, currently 18 states (Arkansas, Illinois, Indiana, Kentucky, Maine, Maryland,
Massachusetts, Mississippi, Montana, New Hampshire, New Mexico, North Dakota, Oklahoma, Oregon,
Pennsylvania, Rhode Island, South Carolina, and Virginia) plus the District of Columbia use a block
system, nine states (Arizona, Georgia, Idaho, Kansas, Michigan, New York, North Carolina, Texas, and
Vermont) use a point system, and ten states (Colorado, Delaware, Iowa, Louisiana, Minnesota, Nevada,
Ohio, Tennessee, Washington, and Wisconsin) operate a hybrid system.120
4) Early Learning Accountability: Current Supports

Because QRIS is based on the child care system—in which providers are given variable reimbursements by the state—the primary method of support through QRIS is often higher reimbursement rates for programs in higher performance tiers, creating a defined spectrum of consequences for each level. States also provide participating programs with training and professional development generally aligned with QRIS requirements, in addition to on-site assistance and quality-improvement grants. The training opportunities made available cover several content areas, ranging from curriculum to language and literacy to environment assessments. The on-site technical assistance helps programs navigate the QRIS process and the implementation of improvement efforts. Meanwhile, improvement grants allow programs to purchase materials, resources, or services to help advance quality-building efforts.

In QRIS, because many of the criteria are inputs or structural issues, the solution to identified deficiencies is frequently straightforward. For example, if a program’s staff is not adequately qualified, the solution to the problem is obvious: hire qualified staff. The same is true of many other common deficiencies, such as offering technical assistance and financial supports to help programs with limited capacity put quality-improvement plans into action. However, the fact that QRIS is able to identify remediable deficiencies does not mean it is able to solve them. In many instances, the differential reimbursement provided is not adequate for programs to improve quality. Moreover, in some instances, the relationship between these discrete improvements and better child outcomes is not clear.

QRIS supports also must be flexible because they reflect the fact that children enrolled in early learning are frequently outside a traditional public school setting. Accordingly, a significant focus of support efforts has been on continuous quality improvement processes.

B) K–12 Accountability

For more than a decade, K–12 accountability has been shaped by the No Child Left Behind Act (NCLB), which reauthorized the Elementary and Secondary Education Act (ESEA)—and, more recently, by the US Department of Education’s waivers from that act. Under NCLB, the federal government provided a framework with performance metrics for schools, measurements, performance tiers, and supports. The waiver process initiated in 2011 by the Department of Education did not change that framework; instead, it allowed for changes in the metrics, measurements, tiers, and supports. As of this writing, 43 states, the District of Columbia, and Puerto Rico have received waivers; the Bureau of Indian Education and two states (Iowa and Wyoming) have requests under review; two states (North Dakota and Vermont) submitted requests that were later withdrawn; and three states (California, Montana, and Nebraska) had not applied. The department has announced plans to extend the waivers for two additional years. An extra one-year extension to implement the teacher-evaluation portion of the waivers was recently approved for Missouri and Nebraska, while the state of Washington’s request was not extended. Washington’s waiver flexibility will end with the 2013–2014 school year, making it the first state to have its waiver flexibility revoked.
A primary goal of the ESEA waiver process was to allow states to redefine their accountability systems and move away from No Child Left Behind’s requirements of reaching 100% proficiency based on state assessments. The waiver process, administered by the Department of Education, defines a set of criteria that states must use for accountability systems in order to obtain a waiver. The criteria are based largely on accountability principles proposed by the Council of Chief State School Officers in its “Principles and Process for State Leadership on Next-Generation Accountability Systems.” These systems are aligned to college- and career-ready standards, focused on student outcomes (including performance, growth, and graduation), and continue NCLB’s requirement of disaggregating data based on student characteristics.

1) K–12 Accountability: Current Performance Metrics

Accountability under ESEA is based on student outcomes, primarily test results and graduation rates. Under NCLB—the first generation of federal accountability to focus on outcomes—the required metrics were proficiency in English-language arts and math, based on assessments measuring student achievement. States were required to achieve 100% proficiency by 2014 and set interim goals for proficiency for each year in between. States were allowed to create their own definitions of proficiency, and the definitions they came up with varied widely. Regardless of definition, however, as of 2014, no state had achieved 100% proficiency.

The limitations of a proficiency-based system quickly became apparent. Because schools were focused on a single performance bar, studies showed that some schools had little incentive to focus resources on children who were already above that bar—or who were so far below it they had little hope of meeting it. Moreover, schools were not given credit for progress they made if a child did not achieve proficiency. For example, if a new 7th-grade student showed up at a school reading at a 2nd-grade level, and in one year the school raised his reading level to 6th-grade, that extraordinary accomplishment would not be recognized by NCLB; it instead required that the school be held accountable for failing to have him reading at a 7th-grade level.

The waivers allow for accountability to include a mixture of proficiency and growth, to give credit where schools helped students improve their performance even if that improvement still left them short of proficiency. Waiver states have added student growth (based on assessment results) into their accountability systems. Importantly, this change represents a step toward the kind of accountability called for in this paper; the addition of growth measures is meant to serve as a proxy for the quality of professional practices, on the theory that outstanding professional practice may lead to growth even when it does not lead to proficiency. While this is still an indirect method of measuring something that could and should be measured directly, it does suggest that the approach recommended here—balance professional-practice metrics with child-outcome metrics—is one that is already emerging in K–12 accountability.

Another important nascent change in the waivers is that they have allowed states to expand accountability to additional subjects beyond English-language arts and math. Some states have used
the opportunity to expand accountability systems to include social studies or science (or both).\textsuperscript{145} Still, a recent analysis showed that only 18 of the 42 waiver states have incorporated additional measures that go beyond traditional NCLB-era measures.\textsuperscript{146} Here too there is room for greater movement in this direction, but the waivers represent a first step away from NCLB’s limited focus on English-language arts and math.

Just as early learning metrics can differ depending on the age of the children served, so too does K–12 accountability use different metrics for different ages. Within K–12, the waiver process contemplates different methods of rating elementary schools and high schools, with graduation rate a key component of high school accountability.\textsuperscript{147} As with NCLB, graduation rates are an element of high school accountability, although the waiver process has reduced their significance.\textsuperscript{148}

\section*{2) K–12 Accountability: Current Performance Measurements}

While the metrics in K–12 accountability may have changed, the measurements have not: standardized tests remain the dominant measurement in K–12 accountability. The waivers have allowed states to use standardized test scores from subjects other than English-language arts and math, but the ESEA waivers do not, by and large, represent a major advance in the use of accountability measures other than test scores.\textsuperscript{149} States have gone beyond raw proficiency scores to develop multi-indicator indices that provide a more well-rounded picture of school performance,\textsuperscript{150} but those indices are still based on standardized test scores.

The standardized tests at the heart of K–12 accountability are rapidly changing. The widespread adoption of Common Core standards has pushed states to update and revise the standards-based accountability assessments they have in place. The Department of Education has supported two consortia, known as the Partnership for Assessment of Readiness for Colleges and Careers (PARCC) and SMARTER Balanced,\textsuperscript{151} to develop new Common Core-aligned assessments; other test vendors have also worked on developing new Common Core-aligned accountability assessments.\textsuperscript{152} In many states, these new assessments are scheduled to roll out over the course of the next few years.\textsuperscript{153} Some states have decided to take a gradual approach to full implementation of these assessments, while others have chosen to withdraw from PARCC and SMARTER Balanced and will use other assessments.\textsuperscript{154}

As noted above, at the high school level, graduation rates are also an important part of accountability. The process of calculating graduation rates has been standardized across states.\textsuperscript{155}

\section*{3) K–12 Accountability: Current Performance Tiers}

While federal and state accountability measures apply to and have consequences for school districts and schools, tiering efforts have focused primarily on schools. Waiver accountability requires states to designate “priority schools” and “focus schools” whose performances need improving, as well as high-performing “reward schools.”
• **Priority schools:** Priority schools are among the lowest performing in the state. The total number of priority schools must be at least 5% of all Title I schools statewide.\(^{156}\) (The Title I program provides eligible schools serving a high at-risk population with financial assistance to meet the needs of their students.\(^{157}\)) A priority school is defined as one of the following:

  • A school among the bottom 5% of Title I schools in the state based on school-wide academic achievement scores\(^ {158}\)
  • A Title I or Title I-eligible high school with a graduation rate below 60% consistently over a period of years
  • A Tier I or Tier II school currently implementing a School Improvement Grant (SIG) intervention model. Tier I and Tier II SIG schools are considered by the state to be persistently low achieving.\(^ {159}\)

• **Focus schools:** Focus schools are Title I schools with significant achievement gaps. Focus schools must include at least 10% of Title I schools statewide. A focus school is defined as one of the following:

  • A school with the largest achievement gaps or difference in graduation rates between the highest- and lowest-performing subgroup of students
  • A school with subgroups of students with low academic performance or poor graduation rates\(^ {160}\)

• **Reward schools:** Reward schools represent Title I schools with a track record of top performance or high progress. A reward school is defined as one of the following:

  • A school that has been identified as having the highest absolute performance for all students and also subgroups on statewide assessments, and at the high school level is among Title I schools with exemplary graduation rates
  • A school that is among the 10% of Title I schools showing the most progress in improving academic achievement gaps among all students and subgroups and also demonstrated increased high school graduation rates\(^ {161}\)

Waivers have significantly changed the metrics used for tiering under state accountability systems, moving the focus from absolute performance (whether or not a school met a state-defined standard) to relative performance.\(^ {162}\) As noted above, states have also used metrics other than test scores and graduation rates to identify priority and focus schools, although test scores continue to be the primary metric.\(^ {163}\)

At the state level, an increasing number of tiering systems are using letter grades.\(^ {164}\) With its A+ Plan for Education, Florida became the first state to institute an A–F grading accountability system under Governor Jeb Bush as part of the push for standards-based reform. The grading of schools on an A–F scale based on
student achievement sought to spur school improvement in a manner that rewarded academic success for schools as well as put into place sanctions and forms of assistance for low-performing “D” and “F” schools. Over the years, the letter-grading scale has been used as a model for 15 other state initiatives, despite mixed reviews. It is worth noting that even the states using letter grades have different criteria for sorting schools into those tiers, including utilizing measures on student learning growth, high school graduation rates, college and career readiness, and dual enrollment; thus, the use of letter grades creates a common tiering and labeling approach but not a common set of metrics.

4) K–12 Accountability: Current Supports

Because K–12 accountability is based on outcomes, the supports necessary to achieve those outcomes are often difficult to pinpoint. In general accountability metrics are used to make an initial categorization of schools, but the actual nature of support to the school must be based on a more individualized and targeted diagnostic review. The ESEA waiver process represents another stage in the evolution of thinking about school improvement, anticipating that states will identify priority and focus schools and design interventions that meet the needs of schools in each of those categories. Both prior to and after the implementation of waivers, state lists of the lowest-performing schools subject to supports and interventions include a disproportionate number of schools with significant percentages of students in poverty.

A critical aspect of K–12 supports is that they do not generally flow directly to schools—they flow to school districts. Generally, efforts by school districts to effectively implement interventions and distribute resources to low-performing schools are limited by lack of capacity, integration and alignment, and buy-in to produce school success. Additionally, district-implemented improvement strategies can be disconnected from the daily activities of schools; without adequate capacity, initiatives to aid schools often fall short of making significant gains and meeting the needs of schools and their students.

It is also important that much of the conversation about the consequences of K–12 accountability has focused less on support than on negative consequences for schools that fail to perform. The last few years have seen a sea change in how districts and schools approach school turnaround, driven in large part by federal School Improvement Grants that force schools to use one of four turnaround models. These consequences are in many instances interventions aimed at changing the oversight of the school—or shutting it down. Consistent with this approach, ESEA waiver accountability requires significant consequences for schools that are among the lowest performing, including closure or major governance changes. These approaches are new enough that no definitive long-term study of their impacts has been possible, although states have struggled with their implementation.

At the K–12 level, accountability systems have historically offered few positive incentives for higher-level achievement. The ESEA waiver process attempts to address this issue by requiring states to identify reward schools that will receive benefits for their high performance or high progress. While limited analysis exists on the effectiveness of these rewards, the offerings given to schools range from monetary rewards to recognition certificates, and they often vary by state.
III) Creating the Next Generation of Education Accountability

To move from the existing accountability systems described in section II to the proposed accountability system described in section I will require substantive changes to existing accountability systems and likely process changes at the state level. This section summarizes what those substantive changes should look like, with some broad recommendations for processes that states can use to achieve change.

A) The Changes That Need to Be Made

In each key element of the accountability system described in section I—metrics, measurements, tiers, and supports—changes need to be made at both the birth-to-five and K–12 levels. In all areas, our view is that the needed changes will not only improve performance compared to existing accountability systems in each sector but will help create a more seamless continuum of accountability. This subsection summarizes our proposed changes for each of those elements.

1) The Changes That Need to Be Made: Performance Metrics

Movement is already under way toward accountability systems that measure both professional practice and child outcomes.

• Early learning accountability is increasing its focus on metrics that measure professional practice, and K–12 accountability is adding growth measures—an indirect measure of professional practice.

• Early learning accountability is more explicitly focusing on metrics that are linked directly to child outcomes, although for the most part it still does not measure child outcomes directly. K–12 accountability has focused for many years on child outcomes.

This shows that the approach advocated here is one that is already gaining traction. There are several challenges, of course.

First, the metrics used for professional practice are either insufficiently precise and focused (in early learning) or indirect (in K–12). In early learning, many states are already in the process of pursuing more precise and focused metrics, and the work under way is likely to yield useful lessons for next-generation systems. In K–12, the push to use growth measures represents a step in the right direction, but a fundamental change is still needed to use direct measures of professional practice. The 5E are a research-based resource that can be used to develop these metrics, but more work is clearly needed to develop professional-practice metrics that could be integrated directly into state accountability systems—and that recognize differences in practice across age spans.

Second, the metrics for child outcomes need substantial work. In early learning, the child metrics that are appropriate for use in school accountability tend not to focus directly on standards-based learning—for example, health outcomes. This challenge is discussed further under “Performance Measurements” on
page 26. In K–12, an emerging trend is to broaden the metrics used to go beyond child outcomes on English-language arts and math tests, but even here the trend is primarily limited to standardized assessments in other academic subjects. Broader thinking may be needed to identify child outcomes that more accurately capture what we want from our education system. Moreover, the current trend toward measuring growth should continue, as child growth is likely to be a better measure of school performance than proficiency.

**PERFORMANCE METRICS**

<table>
<thead>
<tr>
<th>Current Early Learning</th>
<th>Proposed Early Learning and K-12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focused on inputs and structural criteria. Not directly tied to child outcomes.</td>
<td>A mix of metrics balancing child outcomes with professional practice.</td>
</tr>
<tr>
<td><strong>Current K-12</strong></td>
<td><strong>Requires rethinking definitions of child outcomes and professional practices, and making dramatic shifts in metrics for all age spans.</strong></td>
</tr>
<tr>
<td>Student outcomes. Not directly tied to practices, processes, or instruction.</td>
<td>Child outcomes will vary across age spans. Balance between child outcomes &amp; professional practice will vary across age spans.</td>
</tr>
</tbody>
</table>

A potential strategy in this area, particularly at higher grade levels, is to leverage some of the rubrics states have developed for planning school improvement and turnaround efforts. In some instances, states have used test scores to identify which schools will receive intensive supports but then developed diagnostic tools to determine what intensive supports or interventions are needed. Some of the elements articulated in those diagnostic tools may speak to one of the 5E, or represent a key value of school improvement of the state that would be worth elevating to an accountability metric.

One other legal requirement that is important to the use of child-outcome metrics is that NCLB accountability required the reporting of child outcome by subgroup (race, income, special education status, and limited English language proficiency), which has helped identify racial achievement gaps in many communities. The ESEA waiver process has changed the approach to subgroups, with states combining racial subgroups together for reporting purposes. The impact of these changes is unclear. While this issue is not explored in depth here, it is important to note that even in accountability systems remodeled in the way proposed here, child-outcome metrics could be reported by subgroup, and differences in subgroup scores could be used for school accountability purposes.

To date, federal law has discouraged states from developing a thoughtful mix of child outcome and professional practice metrics. Accountability systems should include child-outcome measures to ensure a continued focus on achieving results but should do so in a manner that is broadly inclusive of the many...
outcomes we collectively want for children. Because of the importance of age-appropriate metrics, the mix of metrics should change for different age groups, with some metrics cutting across ages and others (like high school graduation rates) used where appropriate. The balance between professional-practice metrics and child-outcome metrics may also vary across age groups. While it is possible to identify principles that should inform state action in this area, it must be acknowledged that a great deal of work would be needed at the state level to give these principles real-world effect.

2) The Changes That Need to Be Made: Performance Measurements

In both early learning and K-12, there is a history of using accountability measurements that are relatively cheap at scale. To some degree, the impacts that accountability systems have had so far reflect the fact that state systems have gotten what they paid for. If the theory of action is that improved professional practice will lead to improved child outcomes—and the research is quite clear that it does—then changing adult practice may require us to measure that practice directly rather than indirectly. But we must continue to measure child outcomes and must acknowledge that by measuring a limited set of child outcomes we may influence adult behaviors in ways that are detrimental to achieving a broader set of child outcomes that we actually care about.

PERFORMANCE MEASUREMENTS

Accountability for K-2

Currently there is no real accountability for the early elementary years (K-2); these grades are generally included in elementary schools that are rated based on their performance on tests beginning in 3rd grade. State accountability systems do not generally measure anything relating directly to K-2 performance. A birth-through-high-school approach with a continuum of metrics—including metrics based on professional practice—could create real and unprecedented accountability for K-2. Indeed, states may want to consider giving K-2 its own set of metrics, distinct from what came before and what comes afterward. Improved metrics in K-2 might help drive better supports in those years and in the ones that precede them.
Early learning has recognized the importance of measuring professional practice and begun to integrate it more aggressively into accountability systems.\textsuperscript{187} The same is not true of K–12 at scale, although there are elements of this approach in use—specifically, the review teams that some states use to assess schools for turnaround after those schools have been identified through low test scores.\textsuperscript{188} The approach is certainly not novel in other countries; as discussed in section I, England and the Netherlands use an inspectorate to rate professional practice in schools.\textsuperscript{189} This form of measurement may have greater up-front costs than what states currently spend on accountability, but those costs need not be astronomical—and because the close examination of professional practice is likely to yield more immediately actionable information than test scores, the investment may prove to be worth it.

Significant work is also needed in measuring child outcomes. This is certainly true at the early learning level because the research on developmentally appropriate assessment has shown that high-stakes assessment is not appropriate for children prior to 3rd grade.\textsuperscript{190} Some existing measurements that can be used may include those focused on health outcomes, but that may feel too attenuated for states in their education-accountability systems. But at all ages, it is important that high-stakes assessments not be the only child-outcome measures. Some child-outcome measurements beyond high-stakes assessment are already in use, like graduation rates; thoughtful states may identify more that are appropriate for inclusion in accountability systems.

\textbf{3) The Changes That Need to Be Made: Tiers}

While different metrics should be used for different ages (as they are today), a set of common performance tiers across age spans could improve the quality of public reporting and help reinforce connectivity between early learning and K–12. While not all states use QRIS as an early learning accountability system, the majority do, and these quality-rating systems generally use a tiering approach. At the K–12 level, ESEA waivers demand tiering to a degree by requiring states to designate reward, focus, and priority schools;\textsuperscript{191} states generally go beyond those minimal requirements with additional tiers and increasingly are using A–F grading systems.\textsuperscript{192} These approaches can be consolidated into a common tiering framework, which can help give more consistent information to the public across age spans.

One major current purpose of tiering is to drive consequences by setting incremental targets for improvement tied to increased flexibility and support (or at least the avoidance of negative consequences). In QRIS, this frequently occurs through increased payments to highly-rated programs.\textsuperscript{193} The ESEA waiver requirements are another example of this, as the waivers define a discrete universe of actions states must take to address performance in focus and priority schools.\textsuperscript{194} If states incorporate metrics and measures that actually diagnose school needs, then the need for tiering to define supports is reduced dramatically; the supports should be based on the diagnosis. States may still want to tie positive consequences to achieving higher tiers, but these could be more generalized rewards, as these higher-performing schools are likely to be able to take advantage of additional resources or flexibility.
In a new model, performance tiering would primarily serve a public information purpose: It would be a way for schools to set a clear path to improvement and expectations for performance, and for parents and the public to understand the quality of a school, which can help inform a community’s approach to a school and a parent’s decision on whether to enroll a child there. Indeed, one of the main purposes for the development of QRIS has been to build parent awareness around understanding and selecting quality early learning programs. Some states post a summary of program ratings in an effort to help parents make more-informed decisions and to increase the demand for higher quality early learning programs. Similarly, in K–12, accountability ratings are meant to inform parents of the quality of their school, and the move to A–F accountability has been fueled in part by a desire for school ratings that are easy for parents to understand. Based on existing infrastructure, this public information function will likely play out very differently between early learning and K–12; early learning is a market-based system in which many providers have never had attendance boundaries, whereas in K–12, the nature of school choice can vary dramatically from jurisdiction to jurisdiction. Parents weigh a variety of factors, including quality, in choosing early learning programs, but some recent studies have indicated that key factors related to early learning’s market-based structure—such as affordability, location, and availability of programs—are also closely tied to parents’ decision-making. Regardless of how many viable choices a parent has, transparent information about the quality of the school can help improve the exercise of that choice and inform efforts by parents and the community to improve the quality of their schools.

Schools as the Unit of Accountability

Federal and state systems treat individual schools as an important unit of accountability. While there is a logic to that, it can create disconnects as students progress and interventions at one level bear no relationship to the academic program at other levels. For this reason, some commentators have advocated creating accountability and implementing supports in “clusters” that are vertically aligned across schools and grade levels. K–12 does have district-level accountability, although that does not always force connectivity across grade levels.

School-level accountability raises complex questions of how responsible schools should be for activities that go on elsewhere. For example, should preschools be responsible for the performance of their students in elementary school? In the case of early learning, supporters and critics of increasing funding for early learning use long-term results to justify their positions, but that has not led to the inclusion of long-term metrics in early learning accountability. That is partly because of the intractable problem that by the time those long-term results are available, it is far too late to provide useful information to parents or meaningful supports or consequences for the preschools. But the issue of accountability for later performance can also be applied to elementary schools, middle schools, and high schools; indeed, one potential metric of high school accountability is college remediation rate.

The issue of accountability for later performance can be flipped to raise the question of whether high schools should be given responsibility for the performance of feeder schools, to encourage working with them to improve performance—a principle that could in theory also be applied to middle schools, elementary schools, and even preschools. This approach would raise enormous challenges of governance and authority. To date, systems have generally shied away from attempting to create accountability across administrative units, but these approaches may begin to emerge in the coming years.
There is no accepted state-of-the-art formula for setting performance tiers, and states will need to experiment in this area.\textsuperscript{200} As noted in section I, one thing that should be consistent across all age spans is that the highest and lowest rating tiers should be reserved for those schools whose indicators are all telling a consistent story. For example, in K–12, some schools have been given low ratings even though their students have shown impressive growth, which means the school's letter grade may give a misleading impression of the professional practice in the school. An increased focus on the quality of practice and a broader set of metrics should make it easier to reserve the lowest ratings for those schools where student performance is weak and adult professional practice is unlikely to help students improve.\textsuperscript{201} Many states that have used voluntary QRIS will need to reconsider the relationship between the lowest tier of the accountability system and the providers not currently in the accountability system. Beyond the lowest tiers, it will likely take years for states to sort out how they want to differentiate among schools in the middle tiers and how those tiers can be used to convey credible information about performance to the general public.

4) The Changes That Need to Be Made: Supports and Consequences

Existing support systems have generally grouped schools into performance bands (e.g., the lowest-performing 5% of schools) and prescribed a set of supports and consequences for schools in that band. A move to accountability based on diagnosis of professional practice replaces this approach with one that is more surgical—an approach in which the supports are specifically based on the school's identified needs. This approach is impossible in an accountability system based solely on student test scores, which do not provide that kind of diagnostic information.

In early learning, states have already started integrating holistic measures of professional performance into QRIS and are attempting to design supports based on those, particularly in the Early Learning Challenge states.\textsuperscript{205} Those states are required to perform validations of their QRIS over the course
of the grant, which could yield valuable information about whether the new supports actually drive improvement. In K–12, there are also elements of this approach in states that have developed diagnostic tools for school performance. In both early learning and K–12, however, these deeper diagnostics are generally used only for a subset of schools.

In some instances, the lowest-performing schools may show they are unable to improve their performance even with the targeted supports. Existing accountability systems in K–12 have included provisions that when schools do not improve, they undergo governance changes or have their funding pulled. The same is true in early learning, where some states are now restricting government funding to programs that achieve a minimum threshold of quality. This kind of consequence seems appropriate only if the school has been given genuine support to address any identified deficiencies but has still shown itself unable to make changes in a reasonable period of time. While shutting off government funding can be disruptive to families, if a school cannot perform even after receiving assistance then it is appropriate to insist on major changes in school management for the school to continue receiving government funding.

State resources for school improvement are limited, and states’ track record of success is also limited. A benefit of better system-wide diagnostics may be that states can focus their efforts on those schools that need the most help with professional practice and design their supports more effectively as well. This is likely to make state resource use more effective and efficient, a must given how little capacity states have had in this area. States must also be thoughtful about how to leverage local resources (particularly school district resources), which may be a more promising role for states than as large-scale direct service providers.

At the state level, in both early learning and K–12, there has been concern that accountability systems are labeling low performers without leading to any improvement in performance. While targeted supports for low performers may always be needed, the bigger issue is whether states are systemically and sufficiently supporting their educational system to produce the results they want. This has obvious implications in the early learning context, where historically states have funded programs that are not designed to provide standards-based education and that frequently reach only a small percentage of the population. But it also has implications in K–12, where states sometimes make policy decisions that are in significant tension with the goals of their accountability system.

Moreover, it is essential that school improvement not be thought of simply as a school-system responsibility. All schools are in communities, and while state supports may be valuable they are no replacement for community engagement in schools. The role of school districts in this work will vary; districts have different levels of capacity to provide support, and in early learning, many schools are not directly supported by their local district—although ideally they will have productive working relationships with it (or with multiple districts, depending on location and attendance patterns). But this work should go beyond school districts and parents to engage additional community leaders and partners. States should attempt to design support systems that reflect the value of community engagement in school improvement and acknowledge the diverse needs of different schools and communities.
Many of the elements of successful schools are elements that state policy and supports can facilitate, but none of the elements of successful schools can be mandated or imposed by states. For state support systems to be successful, they must be premised on the idea that states can use data to inspire action for improvement, but that in many instances the action for improvement will be taken by actors who are not state employees (or even school employees). But states do have a critical role in defining the policies that lead to strong schools and instructional systems, and in many cases can provide valuable support capacity that allows local leaders to act more effectively. Better metrics, measurements, and performance tiering should help state and local leaders be more thoughtful about their improvement efforts and lead to more efficient resource use on behalf of improving student outcomes.

SUPPORTS

**Current Early Learning**
Typically financial incentives based on child care reimbursement
Professional development and training
On-site technical assistance
Improvement grants

**Current K-12**
Focused on building capacity and changing practices
May be a governance change
May be focused on a specific provider (school) or on a more systemic level (district)

**Proposed Early Learning and K-12**
Supports to schools designed to address issues identified by the observation of professional practice.
*Will require a shift from relying on broad categories of activity to more targeted and customized supports.*

**B) How the Changes Will Happen**

In most states there is no history of dramatic improvement driven by an accountability system at either the early learning or K-12 level. QRIS and ESEA waivers are both new enough that it is far too early to declare success or failure regarding their implementations. Both appear to be based on a sound theory of action grounded in identifying needs and targeting resources to address those needs. But there is no doubt that if they experience widespread success in multiple states it will be a development without precedent in the history of education policy in this country.
While states may have a long way to go to develop successful birth-through-high-school accountability, there are some important steps they can take to get there:

• **Identify key entities and stakeholders.** First, states must decide what entities are necessary to develop a comprehensive birth-through-high-school accountability system. As noted, existing QRIS and ESEA waiver systems are likely in separate agencies. This means that governors and legislative leaders will likely be required to develop a combined system, whether through a state P-20 Council or legislative study group. Different states have different infrastructure in place for addressing issues that cut across early learning and K–12; some states may have no effective entity set up to discuss the creation of a system like this. If there is no existing forum, then a governor or legislature would need to set up a dedicated study committee.

  • While the K–12 accountability system will be housed at the state education agency, governance structures in early learning vary substantially, which means that in most states, two or more agencies will need to be actively involved in developing a more seamless accountability continuum. And regardless of how a state’s executive branch is structured, the dynamic between the executive and legislative branch will also be important to consider in developing unified accountability.

• **Identify technical experts.** While high-level political leadership and stakeholder support would be needed to advance the process of developing a comprehensive accountability system, it is important that in most states this exercise will involve revamping something that already exists, not building something new. That can cut both ways politically—sometimes it is easier to create new systems than modify old ones—but at least states will already have some amount of technical expertise in all the elements of their systems (metrics, measurements, tiering, and supports) because they have been responsible for operating systems. As with the development of the existing systems, the process for revamping accountability will require high-level buy-in and technical expertise to develop new metrics, measurements, tiers, and supports. States may wish to draw on national technical experts (as many have done in the creation of QRIS and waiver systems), or they can start with the confidence that there are already people in their community who know how to make these systems work and can modify them based on an assignment given by the governor and/or legislature. States will also need to examine their data systems to ensure that those systems can support a redesigned accountability system.

• **Commit to investing in system improvement with an understanding that savings are garnered in the long run.** There is no question that the proposed use of deeper diagnostic measures in accountability will increase the expense of administering an accountability system and demand a kind of oversight capacity states may not be used to providing. It will also increase substantially the amount of actionable information states have about school improvement, which could easily lead to pressure for greater spending on school supports. The tradeoff if a change like this is executed should
be that (a) accountability will be a seamless progression rather than disconnected criteria across age spans, and (b) accountability systems will be much more effective at driving improvement, so that professional-practice and child outcomes will improve consistently. In the long run, accountability systems should lower overall costs by improving performance across the board—especially in earlier years—in ways that reduce long-term remedial expenses.\textsuperscript{218} Of course, the benefits of increasing high school graduation rates and long-term performance are well known.\textsuperscript{219} States should only try to design a next-generation accountability system if they are prepared to make the investments necessary to make it work, with the expectation that a well-designed investment will lead to long-term payoffs.

- **Be aware of and primed for changes in standards and practice.** The Early Learning Challenge and the ESEA waiver process have moved state accountability conversations in positive directions. But in all states, as noted, the adoption of a common framework will require real changes in accountability at all levels of the age spectrum. How great the gap is between the existing system and a model system will likely vary from state to state, and within states from age span to age span and grade range to grade range (birth to 3 years, 3 to 5 years, grades K–2 and 3–8, and high school). States must be prepared for the idea that developing a new accountability system may shift how resources are allocated among those groups, and they would need to prepare transition plans to ensure that strong educational practice can continue at all age and grade levels.

These recommendations are all for state leaders, but the federal government also has an essential role to play here. For states to succeed with new accountability systems, it is important that these processes be state-driven and not federally mandated. While the Early Learning Challenge and ESEA waiver processes have helped spark movement toward a better system, neither one really contemplates a complete birth-through-high-school overhaul, and the development of a true next-generation accountability system might bump up against limitations within the challenge and ESEA. There are some ways the federal government could give states the leeway and support they need to update their accountability systems.

- For better or worse, developing next-generation systems will be a multiyear process; by the time a new system is designed and adopted, the first wave of Early Learning Challenge grants will likely be over.\textsuperscript{220} By then Congress may have reauthorized ESEA, but if not, then the emergence of comprehensive plans for birth-to-high-school accountability in states might influence Congress to reauthorize ESEA in a manner that could account for next-generation accountability systems.

- In the short term, if states are interested in developing systems of the kind proposed here, the ESEA waivers, administered by the Department of Education, would likely have to be modified to accommodate them. If the department is interested in allowing states to attempt accountability systems of this kind—or at least willing to indulge the possibility—it could review and update its waiver requirements to ensure it is not preventing states from developing such systems.
In addition to their potential use in ESEA, the accountability principles articulated here could potentially be useful in the reauthorization of important legislation that funds early learning programs, including Head Start and child care. The US Department of Health and Human Services could also integrate some of these principles into its administrative oversight of those programs.
Conclusion

Accountability systems should be based on what we want the accountable parties to be providing. In education, the accountable parties are schools, and they should be providing high-quality professional practice that leads to better child outcomes. The metrics we use to measure progress toward these goals are imperfect, and the supports we provide to meet these goals have been inadequate. There is widespread recognition that we can and must do better, and that just leaving schools to their own devices without additional support is unlikely to substantially improve student outcomes.

Accountability must continue to evolve. The early learning field gained lessons in the administration of early learning programs that led to the development of QRIS, yet research has indicated that the approach currently being used in QRIS may not be measuring the practices that lead to improved child outcomes. The K–12 field learned lessons in NCLB that are now being applied in ESEA waivers accountability. While many strong ideas are embedded in existing state QRIS and ESEA waiver accountability systems, one major goal in the next decade should be to learn lessons from their implementation that help them evolve into even better systems. Some of these lessons can be learned even if the systems are kept separate, but creating a single system could create greater opportunity to learn across age spans. The building blocks of unified accountability are already being put into place, and states can combine them to create a more seamless approach to improving student outcomes.

If states are committed to having accountability systems—and in the short term they clearly are—there is much they can do to make them better. They can set metrics and use measurements that evaluate both professional practice and a broader range of child outcomes, giving schools useful information for improvement and helping adults focus on the full spectrum of outcomes public education is supposed to lead to. They can give parents better information about school performance while targeting resources more effectively to improve low-performing schools. They can gain new information that will allow them to make better system-wide decisions about resource use and implementation. And they can bring together systems that are currently separate, providing a strong reminder that it is the same children who are being served from birth through high school graduation. These unified accountability systems are not guaranteed to produce better professional practices or improved child outcomes, but they would offer a better framework for doing so than the systems in use today.

Join the Conversation at theOunce.org/policyconversations

Ounce Policy Conversations are published by the Ounce of Prevention Fund to stimulate and facilitate discussion on topics of interest in early learning policy. We welcome feedback on this paper, and may address that feedback in subsequent versions of this paper or in other papers. If you have comments or questions, please visit our webpage at theOunce.org/policyconversations or contact Elliot Regenstein at eregenstein@ounceofprevention.org.
Endnotes

1. The authors are grateful to their many colleagues at the Ounce of Prevention Fund who reviewed and commented on drafts of this paper. A special thanks to Ounce colleagues Ann Hanson and Brooke Fisher for their contribution to the sidebar on best practice pedagogy in early learning. The authors are also thankful for input from their external reviewers: Jon-Paul Bianchi, Laura Bornfreund, Justin Cohen, Libby Doggett, Barbara Gebhard, Stacie Goffin, Joan Lombardi, Kelly Maxwell, Laura McSorley, Anne Mitchell, Patrick Murphy, Michael Petrilli, Tom Schultz, Alison Segal, Nancy Shier, Helene Stebbins, Louise Stoney, Albert Wat, Sarah Weber, Marcy Whitebook, Conor Williams, Karen Yarbrough, and Margie Yeager. The views expressed in this paper are those of the authors and do not necessarily represent those of the reviewers.


19. For example, a recent Fordham Institute survey found that many parents value career education, instruction in citizenship and democracy, teaching the ability to work with people of diverse backgrounds, and arts and music instruction. Zeehandelaar, D., and Northern, A. (2013) “What Parents Want: Education Preferences and Trade-offs.” [http://www.edexcellence.net/publications/what-parents-want.html].

20. As Robert Pondiscio, Executive Director of CitizenshipFirst, has said, “I have made no secret of my discontent with the effects of standardized testing, while maintaining support for the principle of accountability.” Pondiscio, R. (March 11, 2014). “Testing, Truth, and Consequences.” Education Week, Bridging Differences blog. [http://blogs.edweek.org/edweek/Bridging-Differences/2014/03/testing_tru.html].
21. The question of whether elementary schools should be responsible for high school graduation rates is a potentially interesting one, noted in the sidebar on page 28. But even if a state wanted to include high school graduation rates in elementary school accountability, the practical challenges—including the time lag in collecting data—would be enormous.


36. For example, states can offer model rubrics for teacher evaluation; states that have done so include Colorado (http://www.cde.state.co.us/educatoreffectiveness/smte-teacher) and Massachusetts (http://www.doe.mass.edu/edeval/model/).


46. Large Countywide and Suburban District Consortium. “21st Century Education Accountability.”


Continuity of care in which children remain with the same teacher over time is also important for establishing the strong relationships that promote attachment between teacher and child, and thus provide the foundation for learning. See Miller, “Secure Attachment.”


52. Ibid.

53. Ibid.

54. Ibid.


58. La Paro et al. “Quality in Kindergarten Classrooms”; Pianta and Hamre, “Conceptualization, Measurement, and Improvement.”


69. Sebring et al. “Essential Supports.”


89. Section 14005 “State Applications” states, “The State will take action to design and implement an integrated
   system of high-quality early learning programs and services.” In “State Fiscal Stabilization Fund,” American
   Recovery and Reinvestment Act of 2009, Sections 14005, 14006, and 14013, Title XIV (Public Law 112-10, February

90. Nine states—California, Delaware, Maryland, Massachusetts, Minnesota, North Carolina, Ohio, Rhode Island,
   and Washington—were first-round winners and granted the full amount they applied for (except for California).
   Five more states—Colorado, Illinois, New Mexico, Oregon, and Wisconsin—were second-round winners and
   received 50% of what they originally applied for. Another round of funding in 2013 provided the states that
   did not initially receive full funding—California and the five second-round winners—with another 25% of their
   supplemental-race-top-early-learning-challenge-grants. Most recently, six states were awarded grants in
   December 2013: Georgia, Kentucky, Michigan, New Jersey, Pennsylvania, and Vermont. US Department of
   Education. (December 19, 2013). “Six States Awarded Race to the Top—Early Learning Challenge (RTT-ELC) Grants

91. As of February 2014, six states do not have a fully implemented QRIS (Missouri requires legislative action, while
   Connecticut, Nebraska, South Dakota, West Virginia, and Wyoming are in the planning stages). QRIS National
   Map%20QRIS%20National%20Learning%20Network%20www.qrisnetwork.org%20%5BRevised%20February%20
   2014%5D.pdf.

   pdf.

93. Ibid., 18–19.

94. For a deeper description of how this works, see Pianta. “Implementing Observation Protocols” and Tout,


96. Ibid., 8–10.

97. Blair, J. “States View Teacher Quality as Pre-K Rating Factor.” (July 11, 2013). Education Week, Vol. 32, No. 35, 31 and
   34. http://www.edweek.org/ew/articles/2013/06/12/35preschool_h32.html; National Center on Child Care Quality
   https://occqrisguide.icfwebservices.com/files/QRIS_Standsrds_Categories.pdf; California Early Learning Quality
   (2012). “Practices for Promoting Young Children's Learning in QRIS Standards.” National Center for Children in

98. Tout. “Child Care Quality Rating System Assessment.”

   Care Quality Rating System Assessment”; Mitchell, A. (2009). “Quality Rating and Improvement Systems as the


105. Ibid.


108. Meisels. “Accountability in Early Childhood”; Lombardi, J. “Additional Comments.” In National Early Childhood Accountability Task Force. “Taking Stock.” For purposes of this paper, we take no position on the propriety and quality of the assessments currently used in 3rd grade and higher to satisfy ESEA’s accountability requirements.

109. “QRISs differ in a key way from K–12 accountability efforts because they focus almost exclusively on inputs into caregiving and caregiving processes rather than on the outcomes of the process, which for K–12 accountability systems are measures of student performance on standardized assessments.” Zellman and Karoly. “Moving to Outcomes,” 10.


122. Tout. “Child Care Quality Rating System Assessment.”

123. Mitchell. “Stair Steps to Quality.”


127. US Department of Education. “ESEA Flexibility Request.”


133. The ESEA waiver process required states to address four principles: College- and Career-Ready Expectations for All Students; State-Developed Differentiated Recognition, Accountability, and Support; Supporting Effective Instruction and Leadership; and Reducing Duplication and Unnecessary Burden. US Department of Education. (June 7, 2012). “ESEA Flexibility.” http://www.ed.gov/esea/flexibility/documents/esea-flexibility-acc.docThis paper addresses only the second principle, State-Developed Differentiated Recognition, Accountability, and Support, which focuses on district- and school-level accountability; it does not address the third principle, Supporting Effective Instruction and Leadership, which focuses heavily on personnel evaluation and individual accountability.


135. Ibid.


140. Guisbond. “NCLB Crashed and Burned.”


142. US Department of Education. “ESEA Flexibility Request.”


144. US Department of Education. “ESEA Flexibility Request.”


147. US Department of Education. “ESEA Flexibility Request.”


156. US Department of Education. “ESEA Flexibility.”


158. US Department of Education. “ESEA Flexibility.”

159. A Tier I school is among the bottom of Title I schools in improvement, corrective action, or restructuring phase, or is a Title I high school that has graduation rates under 60% over several years. Tier II schools are the lowest 5% of secondary schools in the state that are eligible but do not receive Title I funds, or are high schools with consistently low graduation rates. US Department of Education. (November 1, 2010). “Guidance on Fiscal Year 2010 School Improvement Grants. Under 1003(g) of the Elementary and Secondary Education Act.” Office of Elementary and Secondary Education. https://ed.gov/programs/sif/sigguidance11012010.pdf.

160. US Department of Education. “ESEA Flexibility.”

161. Ibid.


Improve?”


175. US Department of Education. “ESEA Flexibility Request.”


177. US Department of Education. “ESEA Flexibility Request.”


179. A call for an improved accountability framework was recently made by a consortium of 15 superintendents representing large suburban districts. The consortium proposed a framework in which learning is measured using multiple metrics that capture a broader range of student learning. Maxwell, L. (February 11, 2014). “Large-District

180. For example, the Oakland, California, Unified School District set comprehensive school-quality metrics for itself that included ensuring thriving students and healthy communities; quality learning experiences for all students; safe, supportive and healthy learning environments; learning communities focused on continuous improvement; meaningful student, family, and community engagement and partnerships; effective school leadership and resource management; and a high quality central office that is in service of quality schools. Its rationale for including these broader metrics was in response to two central concerns: (1) Current accountability systems that focus on high-stakes testing give little incentive for schools to improve and (2) Limited information exists on the effectiveness of district supports and services. The district sought to build a transparent system of accountability that provided a clear understanding of how effective schools and systems work. Oakland Unified School District. (2011). “Community Schools, Thriving Students: A Five-Year Strategic Plan.” http://www.thrivingstudents.org/sites/default/files/Community-Schools-Thriving-Students-Strategic-Plan.pdf.


185. For example, Sam Chaltain argues that we should also measure social and emotional skills and creativity. Chaltain, S. (August 27, 2013). “How Should We Evaluate Our Preschools?” Education Week, Of, By, For: In Search of the Civic Mission of K-12 Schools blog, http://blogs.edweek.org/edweek/civic_mission/2013/08/how_should_we_evaluate_our_preschools.html. Martha Zaslow, Director of the Office for Policy and Communications of the Society for Research in Child Development and a Senior Scholar at Child Trends, noted the importance of thinking beyond child outcomes to include the professionalization of the field, family engagement, and systems building. Zaslow, M. “Thinking of QRIS.” (January 6, 2014). Presentation at National Women's Law Center National and State Child Care Advocates Meeting. For K-12, the Large Countrywide and Suburban District Consortium recommends measuring college- and career-readiness skills, which includes a mastery of deep content knowledge, dual-language ability and fluency, higher-order thinking, and the possession of noncognitive competencies, such as perseverance, self-regulation, and confidence. Large Countywide and Suburban District Consortium. “21st Century Education Accountability.”

186. For example, Achieve Inc. recommends that accountability systems include metrics focused on college and career readiness: obtaining a rigorously defined diploma, assessment scores that indicate college and career readiness, earning college credit while in high school, and minimizing postsecondary remediation rates. Achieve Inc. “Closing the Expectations Gap: 2013 Annual Report,” 27. http://www.achieve.org/files/2013ClosingtheExpectationsGapReport.pdf. These metrics could be integrated into state accountability for high schools as part of an overall continuum, in combination with a focus on professional practice.

187. Twenty states participating in the QRIS require the use of the Environment Rating Scale. Also, as more states develop and establish comprehensive data systems, the use or planned use of the Classroom Assessment and Scoring System (CLASS) into a QRIS or a Race to the Top–Early Learning Challenge application has expanded. Berlin, R. (2012). “Use of CLASS in Quality Rating and Improvement Systems.” http://grisnetwork.org/
resource/2012/use-class%28E2%84%A2-quality-rating-and-improvement-systems; National Center for Child Care Quality Improvement. “QRIS Elements”; Tout. “Child Care Quality Rating System Assessment.”


191. US Department of Education. “ESEA Flexibility Request.”


194. US Department of Education. “ESEA Flexibility Request.”


Achieve Inc. “Closing the Expectations Gap.”


Tennessee) or require QRIS in order to receive public funding (Maine, New Mexico, North Carolina, Oklahoma, and Wisconsin). Vermont’s pre-kindergarten programs are required to be licensed and participate in QRIS, and Maine’s Head Start programs are also required to join the state’s QRIS.


