9

Policy Influences on Ambitious Classroom Instruction, Assessment, and Learning

Laura S. Hamilton, *American Institutes for Research*

José Felipe Martínez, *University of California, Los Angeles*

CONTENTS

INTRODUCTION

Policies and laws enacted at the federal, state, and local levels have influenced school practices in significant and changing ways throughout the history of American public education. Major legislation like the Civil Rights Act of 1964 and Supreme Court cases such as *Brown v. Board of Education of Topeka* have had lasting impacts on education systems and all aspects of school operations, as have countless other policies at all levels of the education system (*Brown v. Board of Education of Topeka*, 1954; Civil Rights Act, 1964; see also Chapter 2 of this volume, "The Struggle to Implement Balanced Assessment Systems: Explanations and Opportunities"). Recent examples of policy impacts include state legislation limiting teaching and discussion of specific topics in schools, as well as local school board decisions around curriculum and textbook adoption. These policy actions influence teaching and learning environments in ways that directly interact with and have the potential to advance or detract from the vision for comprehensive, coherent, and continuous balanced assessment systems described in this volume. Previous chapters have highlighted the wide variety of actors who must be involved in achieving the vision of balanced assessment systems, along with the many conditions that bolster or undermine such a vision. Most of these actors and conditions interact, directly or indirectly, with policies adopted at different levels of the educational and political systems. Therefore, any effort to design and implement a balanced assessment system must grapple with the policy environment and how policy actors engage in that environment.

This chapter aims to build on and update the contributions of numerous other authors who have discussed policy influences on teaching, learning, and assessment, both in the context of balanced assessment systems and more generally (e.g., Darling-Hammond & Adamson, 2010; Marion et al., 2019). The content of this chapter reflects the growing interest in policy to support balanced assessment systems that promote ambitious, high-quality, and equitable learning opportunities for all students. This chapter is structured in three sections. The first provides a brief history of assessment policies and examines their role in supporting teaching and learning. The second explores the limitations of previously enacted policies in promoting ambitious instruction. We consider education policy as a reflection of values and assumptions about the purposes of schooling and discuss how these values and assumptions relate to assessment. The final section discusses implications for designing and implementing policies that promote a balanced approach to assessment and proposes a set of guiding principles and considerations for policy actors. We view federal and state policy makers as the primary but not the only audiences for this chapter.

THE ROLE OF ASSESSMENT POLICY IN SUPPORTING TEACHING AND LEARNING

The term "policy" encompasses a wide variety of laws, regulations, and actions adopted by various institutions, and policies that influence teaching, learning, and assessment are not limited to those enacted specifically to inform the delivery of education. Housing policy, for instance, can contribute to segregation, which can in turn influence students' learning opportunities and outcomes (Brennan et al., 2014; Johnson & Nazaryan, 2019). A comprehensive analysis of how policy influences assessment

is beyond the scope of this chapter; instead, we focus on *assessment policy*, which we define as policy enacted at the federal, state, or local level that mandates, incentivizes, or supports assessing student learning and other outcomes. This definition incorporates a wide assortment of policies including, for example, the accountability requirements under the Elementary and Secondary Education Act (ESEA) of 1965; district requirements for interim or benchmark assessments; how schools use tests to assign students to gifted or accelerated programs; and higher education institutions' use of test scores for admission, placement, or to award credits.

While most K–12 assessments are administered to students in classrooms, choices regarding their content and uses are often made by actors outside the classroom. In this section, we first consider several policy-relevant purposes of assessment and explore assessment policy at the federal and state levels in the United States. We examine additional policy influences and actors, how policy is used to influence teaching, learning, and assessment; and how educators have responded to assessment policy. Finally, we offer contrasting examples from the international literature to illustrate different ways of conceiving the role of assessment in educational systems outside the United States.

Policy-Related Purposes of Assessment

Educational assessments have been used for a variety of purposes in the policy space. Ho (2022) proposed a simple framework for classifying different purposes and uses of tests and assessments. As shown in Figure 9-1, the framework distinguishes high- and low-stakes contexts for using assessment results at the level of individuals or groups.

Although policy discussions often focus on accountability uses in the upper right-hand quadrant, the uses depicted in the other three quadrants are also influenced by policy. Prominent examples in these quadrants include policies related to the National Assessment of Educational Progress (NAEP) or other national or international monitoring tools, state guidelines or district requirements for teachers to administer specific interim assessments in classrooms, and the policies of higher education institutions regarding admissions tests or awarding of Advanced Placement (AP) credit (National

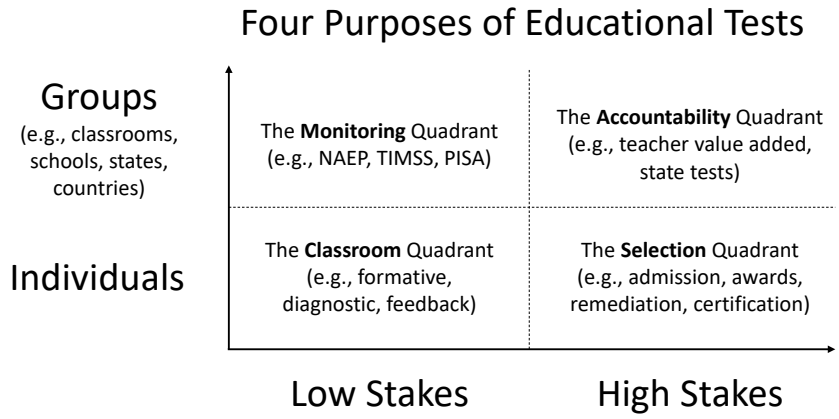

FIGURE 9-1 Purposes and uses of educational assessments.
SOURCE: Ho (2022).

Center for Education Statistics, 2023). In addition, some common examples of assessment policy include features that cut across quadrants in Ho's framework. For instance, school-level report cards might be considered low-stakes in the sense that they are likely not associated with specific rewards or sanctions for students or schools, but they can become high-stakes if they lead to intense public pressure or other—often unintended—consequences. As Hutt and Polikoff (2020) note, "many education policies rely exclusively on the theory that disclosing relevant information to the public about a desired policy outcome—test scores, graduation rates, school climate—will help secure that outcome" (p. 504).

Moreover, "accountability" does not necessarily imply the attachment of consequences to performance. Darling-Hammond (2004) describes five types of accountability: political, legal, bureaucratic, professional, and market. Test scores can be used as part of a *bureaucratic* approach to accountability that aims to motivate improved performance through test-based consequences. This approach, which is not limited to education, is often referred to as performance-based accountability (Stecher et al., 2010). Test scores can also inform *market*-based accountability, particularly in districts or regions that offer public school choice and make scores available to parents to inform that choice (Hamilton & McEachin, 2019). Multiple accountability mechanisms can be present in a specific set of policies, and scores can influence actors differently, even within the same assessment program. For example, in addition to being subject to the formal consequences imposed under current Every Student Succeeds Act (ESSA) accountability provisions (Every Student Succeeds Act, 2015), schools placed in the lowest-scoring category might experience criticism from parents, the press, or other groups, which could be disruptive and demoralizing to school staff and students. But these consequences might also induce beneficial effects that stem from receiving additional resources and support. Schools that are not at risk of falling into the low-performing categories, on the other hand, might primarily experience accountability stemming from public reactions to their assessment results rather than the possibility of formal supports or sanctions. The complexity of accountability-related policies and the unpredictability of actors' responses require a thoughtful approach to evaluating the potential consequences of policies that rely on test scores to inform decisions.

Federal- and State-Level Policy Making[1]

The United States is famously not a national education system, but a collection of 55 separate state and territory systems containing thousands of local subsystems, with enormous variation among them in every respect. Nevertheless, much of the most impactful policy making related to assessment occurs at the federal and state levels. In this section, we describe some highlights of federal and state policy from the past several decades, emphasizing aspects of those initiatives that are relevant to balanced assessment systems. We refer readers to Chapter 7 of this volume, "State Practices and Balanced Assessment Systems," for additional discussion of state assessment policy, and Chapter 2 of this volume, "The Struggle to Implement Balanced Assessment

[1] Throughout this chapter, when we refer to "schools" we are including only K–12 public schools. Private schools are not subject to most of the federal and state policies discussed in this chapter.

Systems: Explanations and Opportunities," for a detailed historical overview of and perspective around these policies.

Large-Scale Assessment as a Legislative Priority

Although federal and state legislation that aims to influence what happens in classrooms is a relatively recent phenomenon, large-scale testing has been a feature of the U.S. public school system for decades. Standardized tests were used as far back as the 1840s to monitor the effectiveness of schools and inform which students were selected for high schools (National Research Council, 1982; Tyack, 1974). Tests of what was called "intelligence" were used for selection and placement into the military beginning in the early 1900s, and enthusiasm for standardized tests as tools for informing student grouping, course placement, and other decisions increased in the ensuing decades (Koretz & Hamilton, 2006). The establishment of NAEP in the 1960s provided the first monitoring tool designed to reflect national trends in student performance over time (Koretz, 1992). The use of assessments for large-scale monitoring gained momentum with ESEA in 1965, which required the administration of standardized tests to gauge the effects of Title I compensatory education provisions (Koretz, 1992). While these tests and systems were not linked to high-stakes accountability decisions, they likely contributed to a propensity among policy makers and the public to view test scores as a key indicator of the outcomes and effects of the education system (Airasian, 1987).

The widely publicized 1983 report *A Nation at Risk: The Imperative for Educational Reform* was a significant contributor to the high-stakes accountability testing movement that launched later that decade (Koretz & Hamilton, 2006; National Commission on Excellence in Education, 1983; Pipho, 1985). Many states had already adopted statewide minimum-competency tests when *A Nation at Risk* was released, and several of them heeded the report's urgent call to measure and improve student learning by attaching financial or other incentives to school-level scores. At the same time, state leaders and other policy makers interpreted the report's findings and recommendations as indicating a need to shift from measuring minimum competency to setting high standards and measuring the attainment of those standards (Koretz, 1992).

A growing emphasis on more rigorous standards and instruction aligned to these standards highlighted the limitations of the multiple-choice format that dominated large-scale testing at that time, as well as the importance of using a broader range of formats to better capture higher-order skills (National Research Council, 2001). Rather than presenting a set of response options from which test-takers must choose, performance assessments consist of tasks that invite test-takers to produce responses in ways that can mirror real-world activity and elicit higher-order thinking skills. Performance assessments may also offer a more meaningful activity for test-takers than a typical standardized test (Darling-Hammond & Adamson, 2010; Stecher, 2010).

The shift away from exclusive reliance on multiple-choice questions in the 1990s was also driven by a growing body of evidence on the influence of testing on teaching and learning. The evidence led to calls for new assessments that would reflect and support high-quality instruction and learning—such as "tests worth teaching to" (Madaus, 1993; Resnick & Resnick, 1992; Shepard, 2021). The 1994 reauthorization of ESEA, called the Improving America's Schools Act (IASA), encouraged states to adopt new standards

that emphasized higher-order skills and required the administration of assessments that would measure students' application of those skills (Improving America's Schools Act, 1994; McDonnell, 2005). In response to IASA and broader societal and economic trends, many states experimented with new assessment formats for their statewide accountability systems. Prominent examples included portfolio assessments developed in Vermont and Kentucky, hands-on and collaborative performance tasks in Maryland and Connecticut, and classroom-based assessments in Washington. Notable multi-state initiatives that incorporated performance tasks included the New Standards Project and the New England Common Assessment Program. The National Research Council (2010) provides a detailed account of these and other similar efforts in the United States in *State Assessment Systems,* including some of the key substantive, technical, and policy aspects of their development and implementation.

Although this wave of innovation in state assessments generated valuable research and laid the groundwork for further technical developments, concerns regarding cost and score reliability led to a renewed reliance on multiple-choice items in state assessment programs since the late 1990s (Mehrens, 2002). The subsequent reauthorization of ESEA, the No Child Left Behind (NCLB) Act of 2001, accelerated this shift by significantly increasing the number of tests that states were required to administer, which in turn also led to the proliferation of associated interim tests (Koretz & Hamilton, 2006; Marion et al., 2019; No Child Left Behind Act, 2001). Even when states were willing and able to support high-quality, performance-based assessments in their NCLB systems, they were typically unable to obtain approval for these assessments. NCLB offered detailed prescriptions for how state tests would be used to monitor proficiency for students across subgroups, along with consequences and interventions for underperforming schools. Particularly noteworthy were the "Adequate Yearly Progress" requirements, through which states set ambitious targets for student performance, ultimately reflecting a goal that 100 percent of students would perform at the proficient level or higher by 2014. As Linn (2003) demonstrated through comparisons with prior performance on U.S. and international assessments, for most schools, these targets were unrealistic.

The lists of federal assessment and accountability requirements became increasingly complex and hard for states to meet—or even monitor accurately. In 2009 the U.S. Department of Education launched the Race to the Top (RTTT) initiative, offering states flexibility and financial incentives to develop new data systems to monitor and promote student learning (U.S. Department of Education, 2009). In 2011, the U.S. Department of Education also offered waivers from NCLB provisions to states that developed or strengthened systems that used student test scores for teacher and school accountability. These policies incentivized states to develop new accountability mechanisms and systems that expanded the uses of available test scores—including, notably, to evaluate teacher performance and effectiveness—but did not require evidence showing validity for these new uses (Baker et al., 2010).

From No Child Left Behind to the Every Student Succeeds Act

As a result of its well-documented technical and policy limitations, the latter part of the NCLB era was marked by a new wave of debate and advocacy around how

assessment and accountability systems might be redesigned to promote college and career readiness and, by extension, more ambitious instruction and learning. Calls increased for "deeper learning," through which learners engage in critical thinking, problem-solving, collaboration, effective communication, and other competencies in academic, social, and emotional learning domains (Hewlett Foundation, 2013). Similarly, educators and organizations around the globe argued that schools should promote "21st-century skills" to prepare young people for success in jobs that would presumably require more complex competencies than in the past (Saavedra & Opfer, 2012).

Policies of this era were centered on the Common Core State Standards (CCSS), which describe essential mathematics and English language arts (ELA) knowledge and skills for college and career readiness (Common Core State Standards Initiative, 2010). However, developers recognized early on that their policy objectives could be derailed if tests were inadequately aligned with the ambitious instructional goals outlined in the standards (Council of Chief State School Officers, 2014). These concerns were borne out in a 2012 evaluation, which found that existing state tests largely failed to capture "deeper learning" (Yuan & Le, 2012). Two assessment consortia grew out of RTTT to produce language arts and mathematics assessments aligned with CCSS for use across multiple states. These are the Partnership for Assessment of Readiness for College and Careers (PARCC),[2] which originally comprised 24 member states and the District of Columbia, and the Smarter Balanced Assessment Consortium (SBAC),[3] consisting of 15 member states. Additional efforts specifically focused on English learners were launched by the WIDA consortium[4] (originally established by Wisconsin, Delaware, and Arkansas, and now comprising 41 states) and ELPA 21[5] (English Language Proficiency Assessment for the 21st Century, used in 10 states). Finally, the Dynamic Learning Maps Alternate Assessment[6] for students with significant disabilities is administered in 21 member states, and the Multi-State Alternate Assessment[7] operates in 13 states and territories at the time of this writing.

Early analyses of PARCC and SBAC found that both incorporated key aspects of deeper learning (Herman & Linn, 2013). However, the widespread state adoption of these assessments, which policy makers originally envisioned, failed to hold. By 2023, SBAC continued to be used in a dozen states and territories—including its open-ended and performance tasks—but a large majority of states have withdrawn from the consortia and adopted their own assessments. Importantly, this does not mean that the states that left SBAC developed these assessments from the ground up. Rather, the need for comparability and efficiency drove many states to adapt or modify consortium tests (e.g., New Jersey's Student Learning Assessment[8] is intended as a shorter version of PARCC), or acquire interim or summative tests that are ostensibly still fundamentally aligned to the CCSS (see Fox et al., 2021; Jochim & McGuinn, 2016; additional details can be found by searching the 50-state comparison archives from the Education Com-

[2] See https://resources.newmeridiancorp.org/research.
[3] See https://smarterbalanced.org.
[4] See https://wida.wisc.edu.
[5] See https://www.elpa21.org.
[6] See https://dynamiclearningmaps.org/dlm-assessments.
[7] See https://www.msaastates.com.
[8] See https://www.nj.gov/education/assessment/resources.

mission of the States[9]). Chapter 2 of this volume, "The Struggle to Implement Balanced Assessment Systems: Explanations and Opportunities," discusses the features of state tests—including their limitations for supporting balanced assessment systems—in greater detail.

ESSA, which replaced NCLB in 2015, maintained a focus on accountability but sought to relieve states, districts, and schools of the most rigid provisions and requirements of its predecessor. ESSA offered greater flexibility around the choice of measures to include for student assessment and the mechanisms of school and teacher accountability (Egalite et al., 2017). ESSA also increased emphasis on the use of school-level growth measures based on four basic indicators: academic achievement, academic growth, graduation, and English proficiency. States can also use the School Quality and Student Success (SQSS) indicator—referred to as the "fifth indicator"—to reflect local priorities and efforts and offer a more holistic picture of student success (Council of Chief State School Officers, 2017).[10]

ESSA ended the requirement for states to use aggregate standardized test scores to evaluate teacher performance, which had been a key provision of the NCLB waivers and RTTT. In 2015, the U.S. Department of Education launched the Innovative Assessment Demonstration Authority (IADA),[11] a novel policy initiative intended to allow states or consortia to apply to develop high-quality, innovative approaches and tools for use in statewide accountability and reporting. Examples of such innovations include competency- and performance-based assessments, as well as interim and instructionally embedded assessments. To date, five states have received approval under IADA to develop new approaches. This experimentation is hindered, however, by some of IADA's requirements. In particular, states must ensure that results—for example, the percentages of students performing at or above the proficient level—are comparable between the innovative assessment and the existing state test, a requirement that is challenging to meet when new assessments are designed to measure key outcomes in new ways (Lyons & Marion, 2016). Another limitation is the assumption that innovative approaches—such as through-year assessment—could simultaneously improve instruction and inform accountability decisions, which has not been borne out by states' experiences (Timberlake, 2023). As of this writing, only three states remain in the IADA program, though the U.S. Department of Education announced in October 2023 that it was expanding the program (Gewertz, 2023). States' experiences with IADA illustrate how policies can both foster and hinder innovation.

Other Policy Influences and Actors

To understand the full scope of federal and state influences on teaching, learning, and assessment, it is important to acknowledge the many factors beyond assessment-related legislation that affect schools. We will not attempt to cover these influences exhaustively, but they include executive orders or non-test-related legislation (e.g.,

[9] See https://www.ecs.org/doctype/50-state-comparison.
[10] We discuss the SQSS indicator in greater detail in the section titled "Educating and Assessing the Whole Learner."
[11] See https://www2.ed.gov/admins/lead/account/iada/index.html.

recent state-level prohibitions regarding the teaching of critical race theory, social and emotional learning [SEL], or other topics), judicial decisions (e.g., Supreme Court decisions regarding affirmative action, which could affect the use of admissions tests), and even prominent tests like NAEP, which despite being intended for monitoring purposes, have influenced the public debate about what it means to be "proficient" (Loveless, 2016; National Academies of Sciences, Engineering, and Medicine, 2017).

Although much of the policy influence on assessment stems from government action at the state and federal levels, it is important to recognize that other levels of governance and institutions can also exert important influence on assessment policy. Chapter 6 of this volume, "District and School Practices and Assessments to Support a Learning-Centered Vision," provides a thorough discussion of how school districts and other local education agencies (LEAs) influence assessment, and Chapter 7 of this volume, "State Practices and Balanced Assessment Systems," explores how LEAs engage with state-level actors to shape decisions about assessment policy and practice. Chapter 8 of this volume, "Developing, Implementing, and Institutionalizing Complex Educational Innovations: Considerations for Balanced Assessment Systems," further examines these interactions among school-, LEA-, and state-level actors. In this section, we briefly examine a small number of other groups whose actions affect assessment policy. Given the numerous constituencies that schools serve, it is not feasible to offer an exhaustive list of these policy actors. Instead, we describe three key groups whose actions intersect with broader policy initiatives in ways that influence K–12 education, to illustrate how the complexity of assessment policy making in the United States influences efforts to create balanced assessment systems.

Local Governing Bodies

As discussed in Chapter 6 of this volume, "District and School Practices and Assessments to Support a Learning-Centered Vision," and Chapter 7 of this volume, "State Practices and Balanced Assessment Systems," federal and state policy undoubtedly exert a powerful influence over assessments—but in many ways, LEAs are the actors that ultimately determine the design and enactment of balanced assessment systems. Chapter 6 of this volume presents a detailed view of the role of district-level decision makers. In this section, we focus on locally elected school boards, which have primary governance responsibility in most U.S. schools and are thus an important stakeholder group in the assessment policy landscape. This system of local governance dates back to the 1600s and takes a variety of forms depending on state and local context (Kogan, 2022).

School boards can engage in assessment policy making in a few ways. They can exert direct influence over decisions about locally adopted assessment systems through their role in approving spending on materials or programs, including assessment tools and resources—and thus can be prime targets for marketing by the types of vendors discussed below. To carry out their responsibility for evaluating the performance of district leaders, boards set performance metrics—including in some cases, test score metrics—and assess progress against them. This can, in turn, create pressure for district leaders to promote test-focused instructional practices that could lead to some of the

negative consequences of test-based accountability discussed later in this chapter. Of course, school boards could use the power of the purse and their supervisory responsibilities over district leadership to promote whole-child, balanced instruction and assessment approaches prioritized in this volume.

Elected school board members are, by definition, accountable to their constituents. In recent years, there have been numerous examples of contentious interactions between school boards and members of the public on topics such as COVID-19 safety protocols, social studies curricula, and SEL (Kogan, 2022). Board members might also face pressure from parents who are interested in data on their children's performance. Especially for parents of children who have been poorly served by the education system or who need additional learning supports (e.g., children with individualized educational plans [IEPs]), the ability to access seemingly objective data—such as from statewide tests—on children's academic performance might be a priority that conflicts with other goals, such as minimizing the footprint of the state test.

The political accountability that board members face can be a mechanism through which assessment results influence decision making, and this form of accountability was a driver of ESEA legislation, as described by Hutt and Polikoff (2020). Media outlets and vendors have capitalized on the growing public availability of data to create their own reports and ranking systems, which can further exacerbate test-related pressure on board members. Research suggests, however, that test scores typically exert no more than a small influence over school board election results (Kogan et al., 2016). Increasingly, board input to districts is affected more significantly by political partisanship, and it often fails to represent the interests of the student populations that those districts serve (Cohn, 2023; Kogan, 2022). The roles of locally elected boards and other local governance bodies (e.g., charter school governance bodies) are therefore potentially complex factors when it comes to adopting balanced assessment systems.

Higher Education Institutions

Another key group of policy actors is the expansive and diverse institutions of higher education (IHEs). Colleges, universities, and other postsecondary education institutions influence the policies and practices of K–12 schools in a variety of ways. Indeed, Baker (2014) explored the far-reaching effects of growing participation in higher education, along with the increasing power wielded by IHEs, on nearly all aspects of society, including economic mobility, politics, and the definition of concepts such as intelligence and merit. Naturally, these institutions have also exerted substantial influence over K–12 education. Of particular relevance to assessment policy are the uses of test scores for IHE admissions and for awarding credit. A 2019 review of state assessment programs found that half of U.S. states had adopted either—or in some cases both—the SAT or the ACT as a high school accountability test as part of their ESSA plans (Olson, 2019). One rationale for this choice, despite the lack of evidence that either test is aligned with any state's standards, was that offering these exams universally would increase equity of access to selective IHEs—a hypothesis that has been supported by some recent empirical evidence (Hurwitz et al., 2015). As more IHEs drop their admissions testing requirements (Nietzel, 2022), it is unclear whether states will continue to rely on these exams for accountability.

Another related set of IHE policies is the use of scores on AP exams to award credit. Through the AP program, high school students can take courses that are designed to provide college-level content and can also take an end-of-course exam (AP at a Glance, n.d.). In some cases, students can earn college credit through good performance on AP exams (with the measure of "good" varying by institution). As part of the "fifth indicator" in ESSA, states can incorporate measures of college and career readiness including participation in AP courses, AP exams, or other indicators of access to advanced coursework, such as International Baccalaureate (IB) participation (Aspen Institute, 2018). These state-level decisions provide opportunities for states to incentivize school-level offerings that could potentially improve access to and degree completion at IHEs that offer credit for these offerings.

Admissions tests and AP credit are two examples of how IHE policies and practices might influence K–12 assessment policy. The K–12 and IHE sectors can connect in many other ways, and increasing those connections has the potential to benefit students by bringing more coherence to the education pipeline and improving access to higher education for all learners. At the same time, efforts to align the K–12 and IHE sectors should reflect shared priorities regarding what students are expected to learn and what kinds of experiences education systems should provide. Otherwise, there is potential for undesirable consequences—for example, if college admissions tests that are not aligned with state standards are used for high school accountability purposes.

Vendors

A third group of policy actors is the large number of developers of curricula, professional learning resources, and assessments, many of whom market aggressively to educators and education leaders. This marketing is evidenced by, among other things, the advertisements one finds in many education-focused magazines or the newsletters that professional organizations send to their members. Marion and colleagues (2019) discuss "assessment proliferation" (p. 14) resulting from several factors, including growth in commercial interim assessment solutions during the NCLB era and an aggressive, and sometimes misleading, marketing push by vendors—for example, some assessment vendors were quick to claim alignment with CCSS when states were exploring options for new, CCSS-aligned assessments (Faxon-Mills et al., 2013). As Shepard (2021) noted, sellers of interim tests often "hijacked" the phrase "formative assessment" to market products that were designed primarily to serve as test-preparation tools. In the SEL realm, developers of curricula and assessments have advertised widely and flooded the mailboxes of educators and other decision makers with marketing materials, often using phrases such as "evidence-based" in ways that do not align with rigorous research standards (Assessment Work Group, 2019; Grant et al., 2017).

Local governance bodies, IHEs, and vendors are examples of stakeholder groups who engage with K–12 schools in ways that could influence policy adoption or enactment—but this list is not comprehensive. Other non-system actors such as employers, civil rights groups, the press, and academic researchers often engage in activities that have the potential to influence assessment policy. Although there is often no direct, causal link between these groups' actions and the enactment of K–12 assessment policy or school-level assessment practices, any effort to promote widespread balanced assess-

ment systems is likely to be shaped by at least some of these groups. As we discuss later in this chapter, active and ongoing engagement with all relevant stakeholder groups can help promote a more coherent and less chaotic set of policies and supports for balanced assessment.

How Policies Are Designed to Shape Instruction, Learning, and Assessment

The large number and variety of policy actors discussed in the previous section, along with the many educational, political, and economic factors that influence them, highlights the complexity of understanding how policy can affect what happens at the school and classroom levels. Assessment policy can influence practice through a variety of mechanisms. Below, we summarize some of the major ways that assessment policy influences decisions at the state or local levels:

- *Informing or constraining curriculum decisions*. Although state accountability tests were not primarily intended to change curriculum, the research reviewed in the next section makes it clear that many of these tests have had that effect, leading to shifts in emphasis on different academic subjects and topics or activities within subjects.
- *Determining the features of tests and test administration*. NCLB dramatically increased the required number of state-administered tests. Moreover, by emphasizing coverage of grade-level standards, NCLB led many states to abandon innovative assessment formats and rely instead on multiple-choice or other item types that could be administered quickly and scored inexpensively. Federal legislation also required states to set and report proficiency levels.
- *Allocating financial resources related to testing*. In addition to the large number of required tests, states received limited funding to develop, administer, and score statewide assessments. These limits placed important constraints on the opportunities to adopt assessment approaches aligned with deeper learning—like human-scored performance assessments—and instead incentivized the adoption of inexpensive closed formats.
- *Specifying uses of test scores.* Policy can also mandate how test scores are used. NCLB accountability provisions emphasized using state test scores to rate schools and districts, and in turn, influenced the allocation of funding and interventions. RTTT and related initiatives went even further and advocated the use of scores to evaluate individual teachers. Meanwhile, many schools and districts have adopted local policies like using test scores to determine grade promotion (National Conference of State Legislatures, 2019). As noted earlier, while it is possible to use a test for multiple different purposes, each of these purposes must be supported with sufficient evidence of validity.
- *Incentivizing continuity*. Policies can also influence test development and score use indirectly through incentives or requirements to maintain comparability with existing tests and scales. For example, while IADA ostensibly seeks to encourage innovation in assessment design, it specifies that new assessments need to

produce scores comparable to existing tests, which greatly limits opportunities for innovation in practice.

Many of these requirements not only constrain local decision making regarding what, when, and how to assess student achievement but can also limit opportunities for innovation and affect how educators and others respond to assessment policies. Moreover, policies reflect policy makers' views on the purposes of schooling and can therefore influence the views of other actors. For example, accountability metrics that emphasize the percentage of students performing at or above the proficient level in mathematics and ELA implicitly suggest that schools should prioritize getting students to perform at a particular level in these two subjects while downplaying schools' contributions to more advanced learning in these subjects and to student performance in other subjects.

Yet, it is worth noting that despite widespread concerns about the assessment and accountability provisions in NCLB and ESSA, these policies were intended to help identify the need for additional or improved inputs to help ensure that students would achieve the desired outcomes. Both pieces of legislation were motivated by persistent disparities in achievement across racial/ethnic and socioeconomic groups, and the requirements for annual testing of every student and public reporting of scores at the subgroup level reflect this motivation. The persistence of these disparities suggests the intended outcomes and the theory of action that motivated NCLB and ESSA have yet to fully materialize. This does not necessarily mean that the entire theory of action is flawed. Some elements could continue to play important roles in a more balanced assessment system. For instance, high-quality statewide assessments can help set expectations for student performance, support large-scale monitoring of systems, identify areas in need of improvement, and inform resource allocation.

Research on Educator Responses to Assessment Policy

The day-to-day work of classroom teachers is arguably the most important factor in determining how state and federal assessment policies influence student learning experiences and outcomes. Decades of research show that teachers are also affected by decisions that school and district leaders make in response to those policies. The large number of policy actors and variability in goals and beliefs both among and within groups highlights the complexity of understanding not only how assessment policy gets made, but also the various mechanisms through which it can influence practice (for reviews of this literature see Faxon-Mills et al., 2013; Hamilton et al., 2012; Jennings & Sohn, 2014).

A concise way of summarizing research on how assessment policy affects teaching is the well-known idea that "what you test is what you get," particularly when high stakes are attached to test scores (Koretz & Hamilton, 2006). A 2013 review by Faxon-Mills and colleagues describes potential changes in curriculum content and emphasis, pedagogical activities, and teacher-student interactions that could result from assessment policy. Within each of these three broad categories, changes could be beneficial, harmful, or neutral, depending on the features of the assessments and policies associated with them. Of particular relevance to this volume is how assessment policy has influenced

teachers' emphasis on ambitious instruction, which comprises both curriculum and pedagogy. Numerous studies have found that high-stakes multiple-choice or short-answer tests used for accountability typically lead teachers to increase time devoted to teaching basic skills and facts (Gallagher & Smith, 2000; Jones et al., 1999; Shepard & Dougherty, 1991). By contrast, assessments designed to measure more complex outcomes, such as the Vermont portfolio program and the Maryland School Performance Assessment Program, are typically accompanied by increased instructional emphasis on higher-order thinking, sophisticated writing, and complex problem-solving (Fuchs et al., 1999; Koretz et al., 1994, 1996; Lane et al., 2002). However, more complex assessments are not always associated with their intended effects; research also suggests that under high-stakes conditions, educators often resort to less ambitious instructional strategies that are intended to raise test scores—for example, "rubric-driven" instruction designed to maximize score gains rather than promote more generalizable skill development (Stecher & Mitchell, 1995).

Educators' responses to assessment policy are influenced by many factors, only some of which are under the direct control of policy makers. Faxon-Mills et al. (2013) identified five categories of conditions that influence educators' responses: (1) features of the testing programs, including the tests themselves and how scores are used; (2) the specific accountability provisions, including stakes attached to scores and metrics used to inform accountability decisions; (3) educators' beliefs, knowledge, and prior experiences; (4) characteristics of schools and students, including prior school performance; and (5) district- and school-level policies, including those around curriculum and professional learning opportunities for teachers.

Like teachers, district and school leaders also make decisions that affect instruction, including the adoption of instructional materials and mandating the amount of time devoted to specific subjects. Research shows that these decisions are often influenced by accountability pressures and that one common response to these pressures is to increase district support for teaching and learning at the school and classroom levels (Hannaway, 2007; Ladd & Zelli, 2002; Opfer et al., 2008; Rentner et al., 2006). On the other hand, the literature also highlights ways in which the logic model of accountability systems can fall short of producing intended district- and school-level responses. One salient example is recent policies and efforts involving the use of state assessment data: a key finding from the literature is that scores from statewide accountability tests have not proven useful for informing instruction, despite claims made by the authors of federal accountability legislation (Mandinach & Gummer, 2021; Marsh & Farrell, 2015; Marsh et al., 2006).

A more recent study of changes to instruction in response to ESSA-era accountability provides additional evidence on how local conditions, including governance and educator support, can affect educators' responses to assessment policy (Polikoff et al., 2022). Finally, broader societal conditions and issues can also influence how assessment policies are translated and enacted at the school and classroom levels—as exemplified by the widespread attention to "unfinished learning" stemming from in-person education interruptions due to the COVID-19 pandemic, which led to calls to adopt more frequent assessments of student achievement in schools, and in turn influenced marketing by vendors wishing to sell such assessments to districts (Jimenez, 2020).

An International Perspective on Assessment Policy

The U.S. education policy landscape is best conceived not as a fixed structure, but as an intricate mechanism with many moving parts that are being constantly updated and revised. This state of constant change presents challenges—but also frequent opportunities—to reorient policy frameworks and structures. In this context, the experiences of other countries can offer useful case examples and counterfactuals in reimagining U.S. educational policies and structures and moving them toward the kind of balanced assessment systems described in this volume. Despite inescapable differences in size, structure, and cultural and political contexts, comparative analysis can help broaden the field's understanding and vision of what balanced assessment might look like in practice and what types of systems are possible. A detailed review of assessment policy and practice in the international context is beyond the scope of this chapter, but the observation that countries that are seen as high performing vary dramatically in how they conceptualize, implement, and use assessments is an important one (see Faubert, 2009; Organisation for Economic Co-operation and Development, 2013). Finland does notoriously little standardized testing but emphasizes a strong culture of ongoing, systematic classroom assessment, including student self-assessment that can inform instructional decision making for teachers. Japan administers national exams in some grades, but they are only reported at the regional level, and schools otherwise have full autonomy over assessment. New Zealand requires reporting to parents on student progress in relation to the national curriculum but gives schools full discretion to adopt or develop meaningful assessments for this purpose, and notably created a national system of assessment for learning to support school capacity building and teacher literacy in classroom assessment. Interestingly, teachers in the three countries just mentioned often stay with a group of students for multiple grades, enabling formative assessment to provide a more robust evidentiary basis to work with parents to monitor and improve student learning over time. Next, we examine two examples of assessment policy and practice in countries other than the United States and consider the value and implications of these comparisons.

Assessment in the Dutch education system comprises multiple formative and summative components implemented at the school level and guided by national curricula and performance targets (reference levels), all within a comprehensive inspectorate framework. Schools are required by law to monitor and report to parents on student achievement and progress regularly during the school year but have full autonomy to choose both the frequency of assessment and the specific tools used for this purpose—drawing on available offerings from the Central Institute for Test Development and other national test developers (Scheerens et al., 2012). Notably, data reported to parents and back to the broader education system also include evidence from classroom assessments developed by teachers, as well as written and oral reports, homework, and projects embedded in the curriculum. The inspectorate framework integrates these markers of academic progress with other social, emotional, and civic learning outcomes. The process incorporates a wide range of indicators reflecting aspects of instruction, climate, and school management valued in the framework, and considers them in concert with the school governing board on 4-year inspection cycles. All students take a summative test at the end of primary education, which is also selected by schools from approved

lists of assessments aligned to the national curriculum. The results of this test, along with other relevant evidence from classroom assessments and projects are integrated into a portfolio, which can inform school improvement efforts, but also student placement in secondary education (Ministry of Education, Culture, and Science, 2021).

The basic components of this system have been in place since the late 1960s. This longevity has helped develop a robust culture of assessment that values the consistency and comparability of standardized tests but also builds on the strengths of formative classroom assessment to enable instructional improvement. Strengthening the capacity of teachers and schools to use formative assessment to improve student learning has also been an important policy priority (Nusche et al., 2014), and the system also explicitly considers the use of standardized tests and classroom assessments to support the needs of special education and linguistic minority students. Finally, system-level monitoring has been occurring for many years and relies on probability samples on two national assessments and international comparative studies like the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMMS).

Singapore offers equally interesting contrasting scenarios to the United States. The country's education system is comparable in size and diversity to the median state in the United States, serving more than half a million students with three national languages and a significant immigrant population. After accelerated development in the second half of the 20th century, Singapore garnered attention as a leader in promoting ambitious instructional standards, with students typically ranking at the top in international assessments. In Singapore's national curriculum, school-based interim assessments are seen as integral to both teaching and learning, and scores are explicitly incorporated into system-level monitoring. Under the Project Work initiative, students carry out a collaborative interdisciplinary project over an extended period, and a portfolio is used to integrate evidence from different assessments and sources (including written reports and oral presentations) reflecting collaborative problem-solving, critical and creative thinking, and knowledge synthesis across content areas (Quek et al., 2007). Interestingly, this emphasis on formative assessment coexists with a strong central system of high-stakes national exams at the end of Grades 6, 10, and 12 that inform important school choice and placement decisions at the higher levels. Amidst robust debates about the limitations of summative assessments, these exams evolved to incorporate a greater variety of questions as well as open-ended oral and written response formats to better assess the types of authentic thinking skills emphasized in classrooms. Singapore's Ministry of Education strongly emphasizes the Project Work initiative and embedded formative assessments as the key to improving teaching and learning, and the Ministry promotes assessment literacy for teachers in these areas as a top priority through publications, workshops, and other tools and resources for teachers (Ministry of Education, 2017).

As a technical matter, it seems clear that systems like those described above would be expected to create conditions conducive to gathering coherent evidence from multiple assessment sources to support ambitious teaching and learning in the classroom—and it is a fact that these same countries routinely outperform the United States in international comparisons (DeSilver, 2017; U.S. Department of Education, 2018). However, extrapolation to assessment policy and practice in the United States car-

ries important caveats. It is important to note that each system reflects assumptions, priorities, and societal values around the purposes of schooling and that these are not always well aligned with standard policy discussions in the United States. For example, compared with many industrialized countries, included those mentioned above, funding for public schools in the United States is less consistent, and substantial variation can be observed both between and within states (Allegretto et al., 2022; Organisation for Economic Co-operation and Development, 2022). On the other hand, less reliance on standardized tests in lower grades in these countries often coexists with strong individual accountability and high-stakes testing for sorting into high school tracks and admission into higher education. Readers should be careful not to idealize or reify assessment systems in other countries or present them as "settled law." As in the United States, there are many important ongoing conversations and, in some cases, intense policy debates—such as in Germany and Canada, which have significantly redesigned assessment systems in response to perceived declines in educational outcomes in international assessments. Interested readers should refer to reviews by Darling-Hammond and McCloskey (2008) and the Organisation for Economic Co-operation and Development (2013).

UNDERSTANDING AND ADDRESSING POLICY LIMITATIONS

The evidence detailed in this chapter and other chapters of this volume (e.g., Chapter 2, "The Struggle to Implement Balanced Assessment Systems: Explanations and Opportunities") suggests that the implementation of balanced assessment systems that support ambitious instruction is rare in the United States (Conley, 2018). Moreover, policy is an inherently blunt instrument that cannot, by itself, induce the kinds of changes needed to achieve the vision of balanced assessment systems. Finally, even the most well-intentioned policies can produce unintended consequences and/or fail to achieve the ambitious goals of their authors. Understanding the role of policy in arriving at the current state of balanced assessment systems in the United States, and the specific challenges faced in a particular context and place, is critical for reimagining education and assessment policy in ways that could help chart a path forward.

One significant challenge in implementing balanced assessment systems is the complexity of the U.S. education system and the large number of actors whose responses to policy are critical for achieving intended outcomes. These actors include educators, vendors, IHEs, and school boards, among others. There can be substantial variability between and within these groups in terms of perceived purposes of schooling, what goals they expect schools to pursue, the most effective levers and strategies to use, and so forth. As a result of this variability, the perceptions and goals of different groups can be directly at odds, potentially giving rise to disagreement and conflict. In addition, differences in the degree and nature of influence afforded to actors in each of these groups can limit the influence of policy on practice. In particular, the significant local control over public education in most states, along with the autonomy that many school leaders and teachers enjoy, can hinder efforts to enact systemic policies related to curriculum and classroom assessment.

An exhaustive overview of factors that explain the failure of policy to lead to balanced assessments and desired outcomes is beyond the scope of this chapter. Some of

these factors are covered in other chapters of this volume, including political factors such as leadership stability (Chapter 2, "The Struggle to Implement Balanced Assessment Systems: Explanations and Opportunities"), district structures (Chapter 6, "District and School Practices and Assessments to Support a Learning-Centered Vision"), and teacher assessment literacy (Chapter 5, "Assessment Literacy and Professional Learning") among others. In this section, we examine how actors' values and beliefs about the fundamental purposes of schools can contribute to their decisions about assessment policy. We then describe three broad policy goals that, if pursued in coherent ways, have the potential to support high-quality, balanced assessment systems.

Connecting Assessment Policy to the Purposes of Schooling

The features of federal and state legislation summarized earlier in this chapter reflect the primacy of the view that achievement in mathematics and ELA are the main outcomes expected from U.S. schools—and, implicitly, that large-scale standardized assessment is a key mechanism by which policy can support this goal. But, as the authors of this volume have noted, education can contribute to much more than academic achievement in a small number of subjects.

Policy decisions necessarily reflect how policy actors think about the types of adults that schools are expected to produce, and more specifically, what outcomes schools are responsible for promoting. Importantly, while these beliefs can play a critical role in policy development and implementation, they are typically not stated explicitly. Policy making could benefit from more systematic and explicit attention and public debate around stakeholders' views about the purposes of schooling. A recent Aspen Institute report advocates for "a deliberative process, engaging students and educators, families, civic and business leaders, and other stakeholders in answering a profound question: **What do we want to be true about public schools in our state?**" (Aspen Institute Education & Society Program, 2022, p. 3, emphasis in original).

Although most Americans would likely agree that schools should ensure that students develop the foundational academic skills necessary to succeed in later pursuits, there is no consensus on the relative importance of these and other outcomes—such as preparing young people for employment or citizenship. Moreover, definitions of concepts like "citizenship" are highly contested (Rapoport & Yemini, 2020). To be sure, a key consideration associated with determining the purposes of schools is *whose* beliefs and values should influence decisions about what schools should emphasize. Lack of consensus on the purposes of schooling can result in fragmented, poorly aligned policy, and in the disenfranchisement of groups who lack political power (Hernández, 2020). Below, we briefly discuss different views about the purposes of public schools that have informed U.S. educational policy.

Schools as Incubators for Citizens

The U.S. public education system was founded on a mission to prepare youth for citizenship (Mann, 1855; Vinnakota, 2019). Despite significant changes to schools' approaches and responsibilities since their founding, public schools continue to be the

primary institutions responsible for developing citizens and civic actors (Winthrop, 2020). Civic learning is not limited to social studies content and includes the knowledge, skills, and dispositions needed to engage constructively in democratic societies (Vinnakota, 2019). The editors of a 2021 National Academy of Education report, *Educating for Civic Reasoning & Discourse*, expressed this idea concisely, noting that "among the most important goals of public education is to prepare young people to engage in informed civic action predicated on a disposition to grapple with the complexities of social issues and policy responses in a diverse society" (Lee et al., 2021, p. 13).

Current events and trends such as the national conversation about systemic racism following the murder of George Floyd, the lack of trust in expertise that became prominent during the COVID-19 pandemic, and growing political partisanship have raised concerns about how well the United States is educating young people to engage effectively in a diverse, democratic society (Blinkoff et al., 2022). Recent polls also indicate that large percentages of high school graduates in the United States express doubts about the health of democracy and their opportunities to effect change (Harvard Kennedy School Institute of Politics, 2022). Although U.S. teachers express support for civic learning, they also report facing significant challenges in ensuring it remains a critical part of instruction, including pressure to improve scores on state accountability tests in other subjects (Hamilton et al., 2020).

Policy solutions proposed to address these challenges have included the adoption of civic learning standards and accompanying statewide assessments—however, the bulk of federal and state policy action has historically emphasized mathematics, English language arts, and to a lesser degree, science. A promising approach to incorporating civic learning could leverage the increased attention paid to social and emotional learning (SEL) in schools in recent years, given the significant overlap between SEL and civic learning frameworks and competencies, e.g., social perspective-taking and cultural competence (Atwell & Bridgeland, 2019; Hamilton & Doss, 2020; Hamilton & Kaufman, 2022; Schwartz et al., 2022; Vinnakota, 2019).

Schools as Engines of Economic Growth

Many factors that detract from schools' efforts to promote civic learning and other aspects of whole-child education stem from pressures related to the role of schools in producing an educated workforce. Policy debates have frequently prioritized schools' economic purpose, emphasizing the need for schools to produce graduates who have the necessary skills to contribute to society through paid work (Zaber et al., 2019). This view of schools as engines of economic success for both individuals and the nation is evident in federal legislation and related policy initiatives like CCSS. These policies typically extend to or feature assessment of academic achievement as a key lever for improving workforce readiness, with actors including government officials, business groups, and parents frequently arguing that assessments should help ensure high school graduates are "college and career ready."

Of course, readiness for college and careers requires not just academic knowledge and skills. Surveys consistently find that some of the most highly sought-after competencies among employers are communication, teamwork, self-management, and integrity (Bauer-Wolf, 2019). Although employers and others often describe these

competencies using terms such as "soft skills" or "employability skills," these constructs in fact map directly onto widely used SEL frameworks (Yoder et al., 2020). Yet, despite widespread consensus that both academic—particularly foundational literacy and numeracy skills—and SEL competencies are necessary to prepare young people to pursue rewarding careers in a variety of fields, most state policy around assessment has emphasized the former to a much greater degree than the latter.

Reducing the Emphasis on Annual Tests

One promising approach to promoting ambitious instruction is a rebalancing of the actual and perceived importance of various elements of the assessment system—particularly end-of-year standardized tests that inform accountability decisions. Although large-scale summative assessments can help monitor outcomes and identify potential sources of inequity, earlier chapters of this volume and related research make it clear that these tests are not designed to support high-quality, ambitious instruction (National Academies of Sciences, Engineering, and Medicine, 2019). Moreover, the stakes attached to scores on these tests, along with reporting that often emphasizes score gaps without acknowledging disparities in access to resources, has the potential to cause harm even if the policies that dictate how scores are used and reported are well intentioned. A heavy emphasis on test scores also signals a narrow set of purposes for the nation's schools—one that is poorly aligned with the whole-child view described above.

Although ESSA provided states with opportunities to expand their accountability metrics, state mathematics and ELA tests persist in carrying the bulk of the weight in ratings. By emphasizing this narrow set of metrics, these systems signal to educators, students, and the public that (1) these are the most important outcomes for schools to promote, and (2) school improvement efforts should aim to increase scores on those tests. State education agencies (SEAs) might not explicitly urge educators to use these tests to inform practice, but their outsized role in measuring school performance sends an implicit message to all stakeholders about the preeminence of these tests. Reducing their salience would require changing federal law and allowing states to experiment with approaches like matrix sampling or reducing the number of grade levels in which testing is required, along with modifications to rules about the identification of individual schools for specific consequences, including sanctions and labeling.

Some writers have proposed through-course or through-year assessments administered multiple times during the school year as a possible way to support balance that serves both summative and formative uses (Javurek, 2020). Such models do not necessarily provide the evidence needed to inform decisions about instruction and accountability, and they often suffer from limitations associated with coverage, precision, and timeliness. They also represent, as Lorié and Dadey (2023) note, a significant change in how states directly influence school activities during the year. Clear guidance from developers and adopting agencies is needed regarding the intended uses of through-year assessment scores, as well as their technical and practical limitations, if this new type of tool will be able to fulfill its promise of serving both formative and summative purposes (Marion, 2021).

Other chapters of this volume provide more detailed explorations of many of these issues. Chapter 4, "Classroom Activity Systems to Support Ambitious Teaching and

Assessment," describes a framework for organizing assessments more tightly around instruction inside the classroom, and Chapter 5, "Assessment Literacy and Professional Learning," highlights the implications of rich conceptualizations of assessment for expectations about teacher professional competencies—and the guidance and resources needed to support these competencies. In this context, thoughtful policy will be needed to focus systems on classroom assessments without imposing constraints or conditions that detract from their utility in informing instruction. Chapter 6, "District and School Practices and Assessments to Support a Learning-Centered Vision," and Chapter 7, "State Practices and Balanced Assessment Systems," outline the types of assessment-related structures, policies, and resources at the district and state levels that are likely to promote desirable assessment practices and prevent or discourage potentially harmful ones.

Educating and Assessing the Whole Learner

In Chapter 3 of this volume, "Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems," Goldman and Lee summarized several decades of research that calls for an integrated view of learning—how cognitive, social, emotional, and cultural factors mediate how learners acquire new knowledge and skills. This conceptual lens is helpful because it strongly indicates that a coherent set of learning goals is likely required to prepare young people for economic success, engaged citizenship, and rewarding relationships. Moreover, COVID-19-related school closures highlighted the many ways that schools contribute to students' development beyond the purely academic, and the aftereffects continue to reverberate not just in students' academic learning but also their social and emotional skills and well-being (Gross & Hamilton, 2023; Hamilton, 2022). This whole-learner perspective is well aligned with recent discussions about the need for accountability systems to incorporate a broader range of constructs (see National Urban League & UNIDOS, 2022). Although this perspective has largely failed to take hold in the assessment systems adopted by educational institutions, in recent years policy makers and assessment developers have taken some steps toward a more expansive view of learner outcomes.

Perhaps the most noteworthy recent initiative reflecting a whole-learner perspective is the so-called "fifth indicator" in ESSA, also known as the "school quality and student success" or SQSS indicator (Council of Chief State School Officers, 2017). States have responded to this flexibility by adding measures like attendance or college and career readiness (Kostyo et al., 2018). If carefully designed, such measures could support more ambitious instruction and signal an interest in a broader set of desired outcomes for schools. For example, several states' ESSA plans include "college and career ready" indices that reward schools whose students participate in opportunities such as completing advanced coursework (e.g., AP or IB) or receiving industry-recognized credentials (Kostyo et al., 2018).

At the same time, reliance on the ESSA indicators to promote ambitious instruction and assessment at the local level has limitations. First, any large-scale assessments of student outcomes that are added to states' ESSA plans will be subject to the limitations of large-scale achievement tests. For instance, states have not yet adopted assessments of SEL competencies in their ESSA systems (Jordan & Marley, 2018). To date, SEL

assessments lack evidence of validity for use in accountability systems or for other high-stakes purposes, and experts have advised states to refrain from including them in their ESSA accountability systems (Assessment Work Group, 2019; Duckworth & Yeager, 2015; Hamilton, 2022; Hamilton & Schwartz, 2019; Melnick et al., 2017). It is also important to note that weights assigned to these indicators in the overall ESSA ratings are quite small relative to the weights assigned to academic achievement tests (Lyons & Brandt, 2021). ESSA might have opened the door for states to adopt a more whole-child approach to accountability, but so far, state movement in that direction is minimal. It is worth noting that in response to the COVID-19 pandemic, the U.S. Department of Education issued guidance in February 2022 that allowed states to modify their plans and increase the weight of the non-academic measures to some degree for 1 year (U.S. Department of Education, 2022).

There is little consensus on whether and how social, emotional, and civic learning competencies should be prioritized in schools, and assessing such competencies presents significant conceptual and technical challenges. Moreover, despite their strong support for SEL (Hamilton & Doss, 2020), educators are increasingly finding themselves enmeshed in highly politicized debates about SEL within school boards and statehouses (Anderson, 2022). Indeed, some groups have pushed back against SEL, conflating it with terms like equity or critical race theory to generate backlash among parents (Joyce, 2022). Nevertheless, an environment that includes increasing calls from employers to instill "transferable" skills in young people and a resurgence in emphasis on the civic mission of schools offers clear opportunities to consider and enact policies that will support both important goals.

Even if states begin to include assessments of SEL competencies or other indicators that reflect a broader perspective on student learning and the purposes of schooling, these efforts are unlikely to reflect the integrated nature of learning described in Chapter 3 of this volume, "Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems," and elsewhere (e.g., Lee et al., 2021). Research demonstrates that the social, emotional, cultural, and academic aspects of learning are integrated (Aspen Institute & National Commission on Social, Emotional, & Academic Development, 2021). Advances in assessment design—including but not limited to technology-based approaches—offer examples of tasks that integrate these dimensions. For instance, Andrews-Todd and colleagues (2019) developed a technology-based assessment that measures mathematics competencies in the context of a collaborative problem-solving environment. Tools like this one have the potential to support assessment that is aligned with an integrated perspective on learning, but more development and research are needed to enable this approach on a large scale and ensure that it has the intended effects on instruction.

Connecting Outcomes to Inputs Through Opportunity-to-Learn Indicators

Assessment policy often seeks to identify areas of need and inform resource allocation, but this cannot be achieved through outcome measures alone, regardless of their breadth or level of detail. Informed decisions require documenting not just learning *outcomes* but also resources and *opportunities* offered to learners to achieve those out-

comes. As a recent Aspen Institute report noted, "Opportunities to learn–the resources, experiences, and expectations students get access to–enable students to pursue their purpose, develop their agency, and contribute as community members and informed citizens" (Aspen Institute Education & Society Program, 2022, p. 2). The report calls on state leaders to take a strategic approach to collecting and making sense of opportunity-to-learn (OTL) data, including through analyses of disparities among groups, clear and actionable reporting mechanisms, and supports for continuous improvement. Similarly, ESSA opens the door for states to include OTL measures in their accountability systems. By itself, this is unlikely to move the needle toward ambitious, whole-child instruction, but it provides a starting point. Darling-Hammond and Cook-Harvey (2018) describe the need for additional funding, guidance, professional development for educators, and family engagement, along with other policy changes.

OTL indicators can help monitor aspects of the learning environment that contribute to ambitious instruction and assessment. This idea aligns with the discussion in Chapter 1 of this volume, "Reimagining Balanced Assessment Systems: An Introduction," on the importance of effective, safe learning environments and a climate that supports whole-child development, as well as the role that resources such as high-quality curricula or caring teachers play in creating such environments. According to the National School Climate Center, climate refers to

> patterns of students', parents' and school personnel's experience of school life [that] reflects norms, goals, values, interpersonal relationships, teaching and learning practices, and organizational structures … [which] foster[s] youth development and learning necessary for a productive, contributing and satisfying life in a democratic society. (National School Climate Center, 2021)

Darling-Hammond and Cook-Harvey (2018) reviewed evidence from the learning and developmental sciences and noted that "a positive school climate is at the core of a successful educational experience" (p. v). Reflecting this growing consensus, eight states included student climate surveys as part of the "fifth indicator" in their ESSA plans (Kostyo et al., 2018).

The idea of incorporating OTL indicators into accountability systems is not new. McDonnell (1995) reviewed efforts to use OTL indicators as policy instruments in the 1980s and 1990s, including as part of a short-lived push for enacting school delivery standards and accountability provisions associated with these standards. The high-water mark of these policy efforts was the Goals 2000: Educate America Act, which called for standards to assess "the sufficiency or quality of the resources, practices, and conditions necessary at each level of the education system to provide all students with an opportunity to learn the material in voluntary national content standards or State content standards" (Goals 2000: Educate America Act, 1994). Thirty years later, a new wave of interest in this area is best exemplified by a report from the National Academies of Sciences, Engineering, and Medicine titled *Monitoring Educational Equity* that calls for the development of systems to monitor educational equity, along with proposals to develop subject-specific OTL standards in language, mathematics, and the arts, among others (see Leung et al., 2021; National Academies of Sciences, Engineering,

and Medicine, 2019; National Association for Music Education, 2020; National Council of Teachers of English, 2019).

McDonnell (1995) emphasized the value of OTL as a generative concept that can offer a vision of high-quality, equitable educational opportunities. OTL data can also help highlight differences in educational experiences and opportunities afforded to different groups of students and how these might relate to disparities in achievement, both across and within groups. However, McDonnell also highlighted the technical and political challenges limiting its use as a policy instrument, many of which continue to be relevant 30 years after Goals 2000. In particular, the precise definition of OTL can differ across contexts, from narrower binary indicators of curriculum coverage, breadth, or depth; to richer operationalizations involving school and classroom processes, pedagogical approaches and instructional practices, school resources, and a range of other elements of the instructional climate. Thus, the collection, interpretation, and reporting of OTL data and its expected relationship to outcomes is not straightforward and may not be feasible without a significant investment of resources and, where accountability is involved, political capital (McDonnell, 1995).

The distinction between factors under and outside the control of the education system is also a challenging concept, as is the difference between *equality* and *equity* in relation to the allocation of opportunities and resources in schools. As defined in *Monitoring Educational Equity*, equity requires that educational opportunities consider students' needs to counter "the effects of structural disadvantages that disproportionately affect different student groups" (National Academies of Sciences, Engineering, and Medicine, 2020, p. 1). Although there is widespread agreement around broad equity goals like these, McDonnell points out that states typically have little incentive to hold themselves accountable for *opportunities* they provide to students, because this accountability could open the door for legal action from individuals or groups, limit flexibility, or have other unintended consequences. Finally, OTL indicators are susceptible to inflation or corruption, given the reliance on self-reporting, especially where specific incentives are involved for the different actors.

TOWARD ASSESSMENT POLICIES THAT SUPPORT BALANCED ASSESSMENT SYSTEMS

In this final section, we build on the research and lessons from enacting assessment policy to consider how future policy might promote the kind of balanced assessment systems described throughout this volume. Because it is difficult to offer detailed recommendations that are relevant across diverse contexts, we highlight broad guidelines and considerations for those tasked with designing assessment policies that could contribute to high-quality, equitable, balanced assessment systems and effective and appropriate use of data (see Box 9-1). Rather than allowing current conditions to constrain the discussion (and recognizing that others have written specifically about ESSA reauthorization; see Marion et al., 2020), we adopt a broader view that is intended to spur discussion and innovation in the research, policy, and practice communities. Some of these ideas or recommendations could be enacted within the constraints of existing federal legislation, whereas others would require changes to that legislation. We present these ideas with federal and state policy makers as the primary audience but hope

BOX 9-1
Guidelines for Designing Assessment Policies to Support Balanced Assessment Systems

- Adopt an inclusive, collaborative approach to policy design and implementation
- Interrogate the values that underlie policy
- Ensure that state policies are informed by an understanding of local variation
- Reduce the state assessment footprint, prioritizing coherence and measures that will inform improvement
- Embrace technological innovation cautiously and responsibly
- Recognize the limits and risks of assessment policy and provide support for navigating the politics

that this material will also be of interest to other policy actors for informing a more balanced and innovative approach to assessment in the future.

Adopt an inclusive, collaborative approach to policy design and implementation. Future changes aimed at promoting high-quality, equity-oriented, and balanced assessment policy design will require deep engagement from those responsible for implementing the policy and those affected by it. Stakeholder participation in policy design and planning, including educators, families, and young people, can be especially valuable for promoting buy-in and wider adoption while simultaneously helping to advance policies that meet stakeholder needs. Importantly, this inclusive approach requires that stakeholders feel genuinely involved and that their input has been seriously considered. Stakeholder voice and participation have become a desideratum of education-related initiatives, but "voice" alone, without real opportunities to engage and have an impact, is likely to result in disempowerment and missed opportunities to design systems that address the experiences and needs of the most important stakeholders.

Of course, many state and local education agencies across the nation are already working to increase stakeholder engagement, and it will be critical for decision makers to learn from existing innovation and experimentation. Additionally, groups that convene educators and policy makers across states, such as the Council of Chief State School Officers Collaboratives, the Council of the Great City Schools, or the National Governors Association could play a critical role as creators of networks to support dissemination and exchange of ideas and shared problem solving. These groups could exercise more direct influence in advancing engagement and collaboration than is typically possible within the purview of the federal government.

Interrogate the values that underlie policy. Policies are not value neutral. Policy sends signals and influences actions in ways that reflect the values of those empowered to design them. Two particularly salient and related values are *trust* and *transparency*. Those who develop or implement assessment policy should articulate their values and examine how policy design or enactment might reflect these values. For instance, the relationship between the state and school districts embedded in policy might indicate

that the state trusts district leaders to pursue the right goals and make sound decisions. Similarly, reporting requirements can be designed to prioritize transparency not just for student data but for system-level conditions and in ways that respect the needs of all stakeholder groups. One potential benefit of the inclusive approach described in the previous recommendation is that it provides an opportunity for groups to discuss and align on the values they want policies to reflect.

Ensure that state policies are informed by an understanding of local variation. The guidelines outlined in this chapter cannot be considered without understanding that the wide variation in assessment policies and practices across local education agencies reflects the influence of numerous factors, including financial resources and capacity, leaders' priorities, and each community's values and goals for its young people. The recommendations we outline here cannot be implemented effectively without considering these influences, particularly those that reflect extant inequities in resources available to educators across districts. For example, well-funded districts serving high-achieving, affluent students, which can typically attract and retain highly qualified staff, might reasonably propose to prioritize developing classroom assessment capacity in their teaching force while de-emphasizing the use of assessment resources provided by the state. By contrast, districts that serve students from less affluent communities or that struggle to attract and retain highly qualified staff might not have the capacity and resources necessary to develop strong assessment expertise and systems at the classroom level. Leaders of such districts might believe that they have no choice but to rely on guidance and supports provided by state testing and accountability systems.

Assessment policy will naturally reflect differences in values and assumptions across states (e.g., some policies treat districts as fundamentally limited in their capacity to effectively develop and implement assessment policy, while others view districts as the key engines of change). As detailed in Chapter 6 of this volume, "District and School Practices and Assessments to Support a Learning-Centered Vision," districts also have different goals and models for improvement. However, irrespective of differences in assumptions, values, and governance structures, those in charge of designing and implementing assessment policy should be careful not to issue punitive mandates that apply mostly to underfunded schools, or conversely, design assessment policies that in practice are overly ambitious and unrealistic for all but the wealthiest schools and communities.

Reduce the state assessment footprint, prioritizing coherence and measures that will inform improvement. Assessment is unlikely to exert beneficial effects unless systems are in place to convert the data gleaned from them into insights that will inform teaching and learning. Crucially, however, the types of insights that can be derived depend, among other factors, on the granularity and (dis)aggregation of the data. Large-scale assessments, for instance, can produce data to inform broad decisions about resource allocation, but cannot and should not be the primary source of day-to-day instructional guidance. A direct implication of the vision of assessment in this volume is that states should consider how funding structures drive assessment policy and practice, imagining scenarios where funds are diverted from accountability testing and purchasing assessment products and into developing assessment competencies that may be more impactful in the long term (Chapter 5 of this volume, "Assessment

Literacy & Professional Learning"). States can add value to assessment systems by supporting or even compelling certain uses of assessments, provided that, as discussed in the previous recommendation, mandates consider the diversity of needs and resources across districts. Systems must be designed to ensure implementation does not replicate or reinforce existing inequities and that burdens imposed on schools are weighed against the expected benefits of assessment in specific contexts.

Considering context is also important for preventing patchwork policy that can undermine the coherence and value of an assessment system. Importantly, coherence at the system level will require changing the narrative around the conceptualization of "accountability." The term does not need to refer to high-stakes testing, and while ESSA was designed to shift the narrative toward an emphasis on improvement, achieving that vision is still a distant goal. Policies can incorporate different approaches to accountability, as discussed earlier in this chapter and in Darling-Hammond (2004). By promoting a broader notion of who is accountable to whom, for what, and under what circumstances, policy can support rather than hinder innovation and ambitious instruction. It is also crucial to keep in mind that the effects of accountability policies stem not just from what content is measured, but from specific design decisions like whether and how cut scores are used and what information is included in public reports. A persistent example illustrating these concerns is found in district policies that require schools to report results of existing interim or formative assessments intended for use by teachers to inform their own instruction, which can add stakes that inadvertently reduce utility for the original intended purpose.

The benefits of annual, statewide, standardized accountability testing systems need to be considered alongside the opportunity cost of de-emphasizing locally developed or implemented assessments that can more readily inform instructional improvement. The policy context is ripe for (re)considering models that reduce the testing footprint by employing matrix sampling, adaptive testing, and alternating or skipping grades altogether (Marion & Lorié, 2023). Policy makers should consider ways to disrupt traditional interpretations of score gaps—for example, by incorporating evidence on disparities in learning opportunities, additional information about within-group variability, aligning reporting to the needs of stakeholder groups, and anticipating and preventing undesirable uses.

Finally, while scarce funding and instructional time can make it tempting to use single tests for multiple purposes for efficiency, it is important to recall that these measures should be used only for purposes for which sufficient validity evidence is available (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Kane, 2006). This guidance is generally understood to apply equally to large-scale, high-stakes tests and classroom assessments of academic, social, or emotional learning (Hamilton & Schwartz, 2019; Jones et al., 2022). Of particular note, there is a dearth of theoretical and empirical work examining validity claims involving high-stakes accountability systems that use aggregate scores (Chalhoub-Deville, 2016; Marion et al., 2016).

Embrace technological innovation cautiously and responsibly. Although not a focus of the chapters of this volume, we recognize that advances in technology, particularly those related to artificial intelligence (AI), are likely to influence both what

and how large-scale systems are assessed in the coming decades. However, adoption of these innovations must be done responsibly, equitably, and with "humans in the loop" (U.S. Department of Education, 2023, p. 7). The use of these technologies presents both significant risks as well as potential benefits. Governments and organizations have begun releasing guidelines for the safe and responsible use of AI, including the *Blueprint for an AI Bill of Rights*, released by the White House Office of Science and Technology Policy in 2022 (White House Office of Science and Technology Policy, 2022). Assessment policy should reinforce and, where necessary, expand on these guidelines to offer criteria that are specific to assessment. These policy supports will be especially important as vendors increasingly market AI-driven assessment products to educators. To be sure, technological advances hold considerable promise in areas like embedded and adaptive testing, or in deploying scenario-based performance tasks to assess complex competencies such as collaborative problem-solving. We believe that over time, these advances will become a central component in the conceptualization, development, and implementation of balanced assessment systems. However, it is likely (perhaps predictable) that in the coming years, the rhetoric—and business—around AI will run far ahead of the evidence needed for robust, sensible, and effective assessment policy. Decision makers should be skeptical of short-term claims equating AI-driven automation with inexpensive, easy access to universal, personalized tests, and particularly claims that these new tests will automatically be culturally responsive, and thus more equitable and valid—no evidence currently supports such claims.

Recognize the limits and risks of assessment policy and provide support for navigating the politics. We noted the bluntness of policy as a lever for change earlier in this chapter, so perhaps the most important recommendation is to recognize that policy alone will be insufficient to achieve the vision of a balanced assessment system—and that any policy carries potential benefits as well as risks. Measurement and assessment are not the only mechanisms to achieve the goals of the public education system. Moreover, while federal and state policy have an important role in creating conditions conducive to balanced assessment systems, policy that is limited to assessment is unlikely to result in significant changes to teaching and learning. Policy will need to address the role of curriculum, professional development, and related supports for balanced assessment.

Of course, policy and politics are deeply intertwined, and educators often find themselves on the front lines of responding to political pushback related to curriculum, instruction, and resources (Woo et al., 2023). Balanced assessment approaches could get caught up in political firestorms, especially to the extent that they incorporate issues that have become controversial, such as SEL or cultural or ethnic studies (Lampen, 2022; Schwartz, 2021). The rise and fall of the movement for accountability around *Opportunity to Learn Standards* in the 1990s illustrates the complex interplay of values and priorities of different groups that ultimately determines whether and how policy takes hold (McDonnell, 1995).

Clear, multi-way communication and frequent engagement with stakeholders are unlikely to eliminate partisan objections but could increase understanding and acceptance. Even in the absence of resistance related to specific curricular or instructional issues, policy actors face challenges stemming from the fact that substantial improve-

ment can take years rather than weeks or months—a timeframe that might be longer than local or state legislators' terms in office. School and district leaders in particular need to understand how to communicate effectively about the nature of educational change and to resist pressure for rapid results. It is also worth noting that the significant impact of political actors on policy implementation reinforces the need for schools to prepare the next generation of citizens—admittedly a long-term fix, but a critical one.

CONCLUSION

The ambitious vision of teaching, learning, and assessment described throughout this volume will not be attainable without concerted and coordinated efforts on the part of actors at all levels of the education system. Policies enacted at the federal, state, and local levels are needed to provide crucial supports and leverage to promote this type of systems change, but the same policies also carry risks of serious unintended consequences that educators should anticipate and consider explicitly. Moreover, the fast pace of technological and societal changes like the increasing presence of AI-driven tools, and evolving conceptions of equitable teaching and learning, will require frequent revisiting of assessment policies and practices in the coming years. Those who design and enact policy, and those who respond to it, should draw on lessons from the policy successes and challenges described throughout this volume, both in the United States and internationally, while adopting a collaborative approach and engaging in frequent monitoring and updating of policies to help steer the system in a direction that will benefit all learners.

REFERENCES

Airasian, P. W. (1987). State mandated testing and educational reform: Context and consequences. *American Journal of Education*, *95*(3). https://doi.org/10.1086/444312.

Allegretto, S., Garcia, E., & Weiss, E. (2022). *Public education funding in the U.S. needs an overhaul. How a larger federal role would boost equity and shield children from disinvestment during downturns*. Economic Policy Institute https://files.epi.org/uploads/233143.pdf.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf.

Anderson, M. (2022, September 26). *How social-emotional learning became a frontline in the battle against CRT*. *NPR*. https://www.npr.org/2022/09/26/1124082878/how-social-emotional-learning-became-a-frontline-in-the-battle-against-crt.

Andrews-Todd, J., Jackson, G. T., & Kurzum, C. (2019). Collaborative problem solving assessment in an online mathematics task. *ETS Research Report Series*, *2019*(1), 1–7. https://doi.org/10.1002/ets2.12260.

AP at a Glance. (n.d.). *AP Central | College Board*. https://apcentral.collegeboard.org/about-ap/ap-a-glance.

Aspen Institute. (2018). *Increasing access to advanced coursework*. Chiefs for Change, Educational Counsel, and Education First, The Aspen Institute. https://www.aspeninstitute.org/wp-content/uploads/2018/04/ESSA-IncreasingAccesstoAdvancedCoursework.pdf.

Aspen Institute & National Commission on Social, Emotional, & Academic Development. (2021). *From a nation at risk to a nation at hope: Recommendations from the National Commission on Social, Emotional, & Academic Development*. The Aspen Institute. https://eric.ed.gov/?q=source%3A%22Aspen+Institute%22&ff1=locDistrict+of+Columbia&id=ED606337.

Aspen Institute Education & Society Program. (2022). *Opportunity to learn, responsibility to lead*. https://www.aspeninstitute.org/wp-content/uploads/2022/09/Aspen-Ed-Soc-Opportunity-Responibility_v8-single.pdf.

Assessment Work Group. (2019). *Student social and emotional competence assessment: The current state of the field and a vision for its future*. Collaborative for Academic, Social, and Emotional Learning. https://measuringsel.casel.org/wp-content/uploads/2019/09/AWG-State-of-the-Field-Report_2019_Exec_Summary_DIGITAL_Final-1.pdf.

Atwell, M. N., & Bridgeland, J. M. (2019). *Ready to lead: A 2019 update of principals' perspectives on how social and emotional learning can prepare children and transform schools*. Collaborative for Academic, Social, and Emotional Learning. https://files.eric.ed.gov/fulltext/ED602977.pdf.

Baker, D. P. (2014). *The schooled society: The educational transformation of global culture*. Stanford University Press.

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. J., & Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers*. Economic Policy Institute EPI Briefing Paper #278. https://www.epi.org/publication/bp278.

Bauer-Wolf, J. (2019, January 16). *Survey: Employers want "soft skills" from graduates*. *Inside Higher Ed*. https://www.insidehighered.com/quicktakes/2019/01/17/survey-employers-want-soft-skills-graduates.

Blinkoff, E., Scott, M., & Hirsh-Pasek, K. (2022, November 21). *Let's educate tomorrow's voters: Democracy depends on it*. Education Plus Development. https://www.brookings.edu/blog/education-plus-development/2022/11/21/lets-educate-tomorrows-voters-democracy-depends-on-it.

Brennan, M., Reed, P., & Sturtevant, L. A. (2014). *The impacts of affordable housing on education: A research summary*. Center for Housing Policy INSIGHTS from Housing Policy Research. https://nhc.org/wp-content/uploads/2017/03/The-Impacts-of-Affordable-Housing-on-Education-1.pdf.

Brown v. Board of Education of Topeka. 347 US 483. (1954). https://www.oyez.org/cases/1940-1955/347us483.

Chalhoub-Deville, M. (2016). Validity theory: Reform policies, accountability testing, and consequences. *Language Testing*, *33*(4), 453–472. https://doi.org/10.1177/0265532215593312.

Civil Rights Act of 1964. 42 U.S.C. 1971. (1964). https://www.archives.gov/milestone-documents/civil-rights-act.

Cohn, C. A. (2023, July 25). *School boards in Chino and Temecula raise worries about local control*. *EdSource*. https://edsource.org/2023/school-boards-in-chino-and-temecula-raise-worries-about-local-control/694698.

Common Core State Standards Initiative. (2010). *Common Core State Standards*. https://corestandards.org.

Conley, D. T. (2018). *The promise and practice of next generation assessment*. Harvard Education Press.

Council of Chief State School Officers. (2014). *Toolkit for evaluating alignment of instructional and assessment materials to the Common Core State Standards*. https://files.eric.ed.gov/fulltext/ED554588.pdf.

Council of Chief State School Officers. (2017). *Identifying a school quality/student success indicator for ESSA: Requirements and considerations*. https://ccsso.org/sites/default/files/2017-10/CCSSOIdentifying-SchoolQualityStudentSuccessIndicator1242017.pdf.

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record*, *106*(6), 1047–1085. https://doi.org/10.1111/j.1467-9620.2004.00372.x.

Darling-Hammond, L., & Adamson, F. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning*. Stanford Center for Opportunity Policy in Education. https://globaled.gse.harvard.edu/files/geii/files/beyond-basic-skills-role-performance-assessment-achieving-21st-century-standards-learning-report_0.pdf.

Darling-Hammond, L., & Cook-Harvey, C. M. (2018). *Educating the whole child: Improving school climate to support student success*. Learning Policy Institute. https://learningpolicyinstitute.org/media/547/download?inline&file=Educating_Whole_Child_REPORT.pdf.

Darling-Hammond, L., & McCloskey, L. (2008). Assessment for learning around the world what would it mean to be internationally competitive? *Phi Delta Kappan*, *90*(4), 263–272.

DeSilver, D. (2017). *U.S. students' academic achievement still lags that of their peers in many other countries*. Pew Research Center. https://www.pewresearch.org/short-reads/2017/02/15/u-s-students-internationally-math-science.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237–251. https://doi.org/10.3102/0013189x15584327.

Egalite, A. J., Fusarelli, L. D., & Fusarelli, B. C. (2017). Will decentralization affect educational inequity? The Every Student Succeeds Act. *Educational Administration Quarterly*, *53*(5), 757–781. https://doi.org/10.1177/0013161X17735869.

Elementary and Secondary Education Act of 1965. 20 U.S.C. 6301. (1965). https://www.govinfo.gov/content/pkg/COMPS-748/pdf/COMPS-748.pdf.

Every Student Succeeds Act. 20 U.S.C. 6301. (2015). https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf.

Faubert, V. (2009). *School evaluation: Current practices in OECD countries and a literature review*. OECD Education Working Papers, No. 42. OECD Publishing. http://dx.doi.org/10.1787/218816547156.

Faxon-Mills, S., Hamilton, L. S., Rudnick, M., & Stecher, B. M. (2013). *New assessments, better instruction? Designing assessment systems to promote instructional improvement*. RAND Corporation. https://www.rand.org/pubs/research_reports/RR354.html.

Fox, L., Hartog, J., & Larkin, N. (2021). *The reliability of shorter assessments in New Jersey for group-level inferences*. Regional Educational Laboratory Mid-Atlantic and Mathematica. https://ies.ed.gov/ncee/rel/regions/midatlantic/pdf/REL_2021106.pdf.

Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L., & Katzaroff, M. (1999). Mathematics performance assessment in the classroom: Effects on teacher planning and student problem solving. *American Educational Research Journal*, *36*(3), 609–646. https://doi.org/10.3102/00028312036003609.

Gallagher, T., & Smith, A. (2000). *Main report: The effects of the selective system of secondary education in Northern Ireland*. Queen's University Belfast and University of Ulster. https://www.ulster.ac.uk/__data/assets/pdf_file/0020/223409/2000_The_Effects_Of_The_Selective_System_Of_Secondary_Education_In_Northern_Ireland_Main_Report.pdf.

Goals 2000: Educate America Act. 20 U.S.C. 5801. (1994). https://www.govtrack.us/congress/bills/103/hr1804/text/enr.

Grant, S., Hamilton, L. S., Wrabel, S. L., Gomez, C. J., Whitaker, A., Leschitz, J. T., Unlu, F., Chavez-Herrerias, E. R., Baker, G., Barrett, M., Harris, M., & Ramos, A. (2017). *Social and emotional learning interventions under the Every Student Succeeds Act: Evidence review.* RAND Corporation. https://doi.org/10.7249/RR2133.

Gross, B., & Hamilton, L. S. (2023). *Student mental health and well-being: A review of evidence and emerging solutions.* Center on Reinventing Public Education. https://crpe.org/wp-content/uploads/SEL-Report-2023-Final.pdf.

Hamilton, L. S. (2022). Asking the right questions about using non-academic indicators for school accountability. In F. M. Hess & R. J. Martin (Eds.), *Reflections in the future of K–12 assessment and accountability* (pp. 33–45). American Enterprise Institute.

Hamilton, L. S., & Doss, C. J. (2020). *Supports for social and emotional learning in American schools and classrooms: Findings from the American Teacher Panel*. RAND Corporation. https://doi.org/10.7249/RRA397-1.

Hamilton, L. S., & Kaufman, J. H. (2022). Indicators of equitable civic learning in U.S. public schools. *Educational Assessment*, *27*(2), 187–196. https://doi.org/10.1080/10627197.2022.2087623.

Hamilton, L. S., & McEachin, A. (2019). Accountability and school choice. In M. Berends, A. Primus, & M. G. Springer (Eds.), *Handbook of research on school choice* (pp. 287–301). Routledge, Taylor & Francis Group.

Hamilton, L. S., & Schwartz, H. L. (2019). *Get smart about social and emotional learning measurement*. American Enterprise Institute. https://www.aei.org/wp-content/uploads/2019/08/Get-Smart-About-Social-and-Emotional-Learning-Measurement.pdf?x91208.

Hamilton, L. S., Kaufman, J. H., & Hu, L. (2020). *Preparing children and youth for civic life in the era of truth decay: Insights from the American Teacher Panel.* RAND Corporation. https://doi.org/10.7249/RRA112-6.

Hamilton, L. S., Stecher, B. M., and Yuan, K. (2012). Standards-based accountability in the United States: Lessons learned and future directions. *Education Inquiry*, *3*(2), 149–170. https://doi.org/10.3402/edui.v3i2.22025.

Hannaway, J. (2007, November 17). *Unbounding rationality: Politics and policy in a data rich system*. Mistifer Lecture. University Council of Education Administration.

Harvard Kennedy School Institute of Politics. (2022). *Harvard Youth Poll, 43rd edition, Spring 2022*. https://iop.harvard.edu/youth-poll/43rd-edition-spring-2022.

Herman, J., & Linn, R. (2013). *On the road to assessing deeper learning: The status of Smarter Balanced and PARCC assessment consortia*. National Center for Research on Evaluation, Standards, and Student Testing Report 823. https://eric.ed.gov/?id=ED540597.

Hernández, L. E. (2020). *The challenges and opportunities of advancing equity through education policy: A research overview*. Spencer Foundation. https://spencerfoundation.s3.us-east-2.amazonaws.com/store/5c03521f6cadb53faca44c10d770c682.pdf.

Hewlitt Foundation. (2013). Deeper learning competencies. https://hewlett.org/wp-content/uploads/2016/08/Deeper_Learning_Defined__April_2013.pdf.

Ho, A. (2022). *Towards assessment literacy: Questions to ask about educational tests*. Harvard Graduate School of Education. https://scholar.harvard.edu/files/andrewho/files/testquestions.pdf.

Hurwitz, M., Smith, J., Niu, S., & Howell, J. (2015). The Maine question: How is 4-year college enrollment affected by mandatory college entrance exams? *Educational Evaluation and Policy Analysis*, *37*(1), 138–159. https://doi.org/10.3102/0162373714521866.

Hutt, E., & Polikoff, M. S. (2020). Toward a framework for public accountability in education reform. *Educational Researcher*, *49*(7), 503–511. https://doi.org/10.3102/0013189X20931246.

Improving America's Schools Act of 1994. 20 U.S.C. 6301. (1994). https://www.govinfo.gov/content/pkg/COMPS-753/pdf/COMPS-753.pdf.

Javurek, A. (2020). *It's time to rethink state assessment*. NWEA. https://www.nwea.org/resource-center/white-paper/51283/State-Solutions-Through-Year-Whitepaper.pdf.

Jennings, J., & Sohn, H. (2014). Measure for measure: How proficiency-based accountability systems affect inequality in academic achievement. *Sociology of Education*, *87*(2), 125–141. https://doi.org/10.1177/0038040714525787.

Jimenez, L. (2020). *Student assessment during COVID-19*. Center for American Progress. https://www.americanprogress.org/article/student-assessment-covid-19.

Jochim, A., & McGuinn, P. (2016). The politics of the Common Core assessments: Why states are quitting the PARCC and Smarter Balanced testing consortia. *Education Next*, *16*(4), 44–52.

Johnson, R. C., & Nazaryan, A. (2019). *Children of the dream: Why school integration works*. Basic Books.

Jones, M. G., Jones, B. D., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, *81*(3), 199–203.

Jones, S. M., Lesaux, N. K., & Barnes, S. P. (Eds.). (2022). *Measuring noncognitive skills in school settings: Assessments of executive function and social-emotional competencies*. Guilford Press.

Jordan, P. W., & Marley, P. (2018). *How did ESSA's "non-academic" indicator get so academic? FutureEd*. https://www.future-ed.org/how-did-essas-non-academic-indicator-get-so-academic.

Joyce, K. (2022, April 22). What is "social emotional learning"—and how did it become the right's new CRT panic? *Salon*. https://www.salon.com/2022/04/22/what-is-social-emotional-learning--and-how-did-it-become-the-rights-new-crt-panic.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th edition, pp. 17–64). Praeger.

Kogan, V. (2022). Locally elected school boards are failing: Pandemic stress-tested school governance, revealing many flaws. *Education Next*, *22*(3), 8–13.

Kogan, V., Lavertu, S., & Peskowitz, Z. (2016). Do school report cards produce accountability through the ballot box? *Journal of Policy Analysis and Management*, *35*(3), 639–661. https://doi.org/10.1002/pam.21912.

Koretz, D. (1992). State and national assessment. In M. C. Alkin (Ed.), *Encyclopedia of educational research* (6th edition, pp. 1262–1267). American Educational Research Association.

Koretz, D., & Hamilton, L. S. (2006). Testing for accountability in K–12. In R. L. Brennan (Ed.), *Educational measurement* (4th edition, pp. 531–578). Praeger.

Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice*, *13*(3), 5–16. https://doi.org/10.1111/j.1745-3992.1994.tb00443.x.

Koretz, D., Mitchell, K., Barron, S., & Keith, S. (1996). *Final report: Perceived effects of the Maryland School Performance Assessment Program (CSE Technical Report 409)*. National Center for Research on Evaluation, Standards, and Student Testing. https://cresst.org/wp-content/uploads/TECH409.pdf.

Kostyo, S., Cardichon, J., & Darling-Hammond, L. (2018). *Making ESSA's equity promise real: State strategies to close the opportunity gap*. Learning Policy Institute. https://learningpolicyinstitute.org/product/essa-equity-promise-report.

Ladd, H. F., and Zelli, A. (2002). School-based accountability in North Carolina: The responses of school principals. *Educational Administration Quarterly*, *38*(4), 494–529. https://doi.org/10.1177/001316102237670.

Lampen, C. (2022, April 19). What is the conservative beef with "social-emotional learning"? *The Cut*. https://www.thecut.com/2022/04/conservative-backlash-social-emotional-learning.html.

Lane, S., Parke, C. S., & Stone, C. A. (2002). The impact of a state performance-based assessment and accountability program on mathematics instruction and student learning: Evidence from survey data and school performance. *Educational Assessment*, *8*(4), 279–315. https://doi.org/10.1207/S15326977EA0804_1.

Lee, C. D., White, G., & Dong, D. (Eds.). (2021). *Educating for civic reasoning and discourse*. National Academy of Education.

Leung, M., Cardichon, J., Scott, C., & Darling-Hammond, L. (2021). *Inequitable opportunity to learn: Access to advanced mathematics and science courses*. Learning Policy Institute. https://learningpolicyinstitute.org/sites/default/files/product-files/CRDC_Course_Access_REPORT.pdf.

Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, *32*(7), 3–13. https://doi.org/10.3102/0013189X032007003.

Lorié, W., & Dadey, N. (2023, April 6). Through-year assessment: A radical renegotiation of the school accountability contract. *Center for Assessment.* https://www.nciea.org/blog/through-year-assessment-a-radical-renegotiation-of-the-school-accountability-contract.

Loveless, T. (2016, June 13). The NAEP proficiency myth. *The Brown Center Chalkboard.* https://www.brookings.edu/blog/brown-center-chalkboard/2016/06/13/the-naep-proficiency-myth.

Lyons, S., & Brandt, C. (2021, March 16). *Letting go of labels in school accountability systems*. Center for Assessment. https://www.nciea.org/blog/letting-go-of-labels-in-school-accountability-systems.

Lyons, S., & Marion, S. (2016). *Comparability options for states applying for the Innovative Assessment and Accountability Demonstration Authority: Comments submitted to the United States Department of Education regarding proposed ESSA regulations.* National Center for the Improvement of Educational Assessment. https://www.nciea.org/wp-content/uploads/2021/11/Comparability-Recommendations-for-Section-1204-Pilots.pdf.

Madaus, G. F. (1993). A national testing system: Manna from above? A historical/technological perspective. *Educational Assessment*, *1*(1), 9–26. https://doi.org/10.1207/s15326977ea0101_2.

Mandinach, E. B., & Gummer, E. S. (2021). *The ethical use of data in education: Promoting responsible policies and practices.* Teachers College Press.

Mann, H. (1855). *Lectures on education*. Ide & Dutton.

Marion, S. (2021, November 30). Collaboratively learning about through year assessments. *Center for Assessment.* https://www.nciea.org/blog/collaboratively-learning-about-through-year-assessments.

Marion, S., & Lorié, W. (2023). *Can we reduce testing in K–12 schools? Yes, but there are tradeoffs*. Large Scale Assessment. https://www.nciea.org/blog/can-we-reduce-testing-in-k-12-schools.

Marion, S., Thompson, J., Evans, C., Martineau, J. & Dadey, N. (2019). *A tricky balance: The challenges and opportunities of balanced systems of assessment*. Paper presented at the Annual Meeting of the National Council on Measurement in Education Toronto, Ontario April 6, 2019. https://www.nciea.org/library/tricky-balance-challenges-and-opportunities-balanced-systems-assessment.

Marion, S., Lyons, S., & D'Brot, J. (2016). *Developing a theory of action to support high quality accountability systems design*. Center for Assessment. https://www.nciea.org/wp-content/uploads/2021/11/Theories-of-Action-for-ESSA_040506_2.pdf.

Marion, S., Domaleski, C., & Brandt, C. (2020). *Assessment and accountability recommendations for the next reauthorization of the Elementary and Secondary Education Act*. Center for Assessment. https://www.nciea.org/wp-content/uploads/2020/01/Center-for-Assessment_ESEA.ReauthorizationReport.pdf.

Marsh, J. A. & Farrell, C. C. (2015). How leaders can support teachers with data-driven decision making: A framework for understanding capacity building. *Educational Management Administration & Leadership*, *43*(2), 269–289. https://doi.org/10.1177/1741143214537229.

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education: Evidence from recent RAND research.* RAND Corporation. https://doi.org/10.7249/OP170.

McDonnell, L. M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis*, *17*(3), 305–322.
McDonnell, L. M. (2005). No Child Left Behind and the federal role in education: Evolution or revolution? *Peabody Journal of Education*, *80*(2), 19–38. https://doi.org/10.1207/s15327930pje8002_2.
Mehrens, W. A. (2002). Consequences of assessment: What is the evidence? In G. Tindal & T. M. Haladyna (Eds.), *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation* (pp. 149–177). Lawrence Erlbaum Associates Publishers.
Melnick, H., Cook-Harvey, C. M., & Darling-Hammond, L. (2017). *Encouraging social and emotional learning in the context of new accountability.* Learning Policy Institute. https://learningpolicyinstitute.org/media/111/download?inline&file=Social_Emotional_Learning_New_Accountability_REPORT.pdf.
Ministry of Education. (2017). *Singapore curriculum philosophy*. https://www.moe.gov.sg/about/singapore-teaching-practice/singapore-curriculum-philosophy.
Ministry of Education, Culture, and Science. (2021). *Inspection framework for the supervision of preschool education and primary education.* https://english.onderwijsinspectie.nl/binaries/onderwijsinspectie_eng/documenten/publications/2021/12/13/inspection-framework-primary-education-2021.
National Academies of Sciences, Engineering, and Medicine. (2017). *Evaluation of the achievement levels for mathematics and reading on the National Assessment of Educational Progress*. The National Academies Press. https://doi.org/10.17226/23409.
National Academies of Sciences, Engineering, and Medicine. (2019). *Monitoring educational equity*. The National Academies Press. https://doi.org/10.17226/25389.
National Academies of Sciences, Engineering, and Medicine. (2020). *Building educational equity indicator systems: A guidebook for states and school districts.* The National Academies Press. https://doi.org/10.17226/25833.
National Association for Music Education. (2020). *National opportunity to learn standards: Facilities, personnel, and curricular resources.* https://nafme.org/wp-content/uploads/2020/08/NAfME-Opportunity-to-Learn-Standards-2020.pdf.
National Center for Education Statistics. (2023). *NAEP: National Assessment of Educational Progress.* https://nces.ed.gov/nationsreportcard.
National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* U.S. Government Printing Office.
National Conference of State Legislatures. (2019). *Third grade reading legislation.* https://www.ncsl.org/research/education/third-grade-reading-legislation.aspx.
National Council of Teachers of English. (2019). *Statement on the opportunity to learn*. https://ncte.org/statement/opptolearnstandards.
National Research Council. (1982). *Ability testing: Uses, consequences, and controversies.* National Academy Press. https://doi.org/10.17226/19562.
National Research Council. (2001). *Knowing what students know: The science and design of educational assessment.* National Academy Press. https://doi.org/10.17226/10019.
National Research Council. (2010). *State assessment systems: Exploring best practices and innovations: Summary of two workshops.* The National Academies Press. https://doi.org/10.17226/13013.
National School Climate Center. (2021). *What is school climate and why is it important?* https://schoolclimate.org/school-climate.
National Urban League and UNIDOS US. (2022). *Education assessment, accountability, & equity.* https://nul.org/sites/default/files/2022-09/Emerging%20Areas%20of%20Agreement%209.22.22.pdf.
Nietzel, M. T. (2022). *More than 80% of four-year colleges won't require standardized tests for fall 2023 admissions*. https://www.forbes.com/sites/michaeltnietzel/2022/11/15/more-than-80-of-four-year-colleges-wont-require-standardized--tests-for-fall-2023-admissions/?sh=4fdd31217fb9.
No Child Left Behind Act of 2001. 20 U.S.C. 6301. (2001). https://www.congress.gov/bill/107th-congress/house-bill/1/text.
Nusche, D., Braun, H., Halász, G., & Santiago, P. (2014). *OECD reviews of evaluation and assessment in education: Netherlands 2014*. Organisation for Economic Co-operation and Development.
Olson, L. (2019). *The new testing landscape: How state assessments are changing under the federal Every Student Succeeds Act*. *FutureEd*. https://www.future-ed.org/wp-content/uploads/2019/09/FutureEdTesting LandscapeReport.pdf.

Opfer, V. D., Henry, G. T., & Mashburn, A. J. (2008). The district effect: Systemic responses to high stakes accountability policies in six southern states. *American Journal of Education*, *114*(2). https://doi.org/10.1086/521242.

Organisation for Economic Co-operation and Development. (2013). *Synergies for better learning: An international perspective on evaluation and assessment*. Organisation for Economic Co-operation and Development. https://www.oecd.org/education/school/Evaluation_and_Assessment_Synthesis_Report.pdf.

Organisation for Economic Co-operation and Development. (2022). *Subnational data: Financial resources invested in education*. https://stats.oecd.org/Index.aspx?DataSetCode=EDU_FIN_SUBNAT.

Pipho, C. (1985). Tracking the reforms, Part 5: Testing—Can it measure the success of the reform movement? *Education Week*, *4*(35), 19.

Polikoff, M. S., Stornaiuolo, A., Flores, N., Desimone, L. M., Pak, K., Fuchs, L. S., Porter, A. C., Smith, T. M., Fuchs, D., Garet, M. S., Song, M., & Nichols, T. P. (2022). *The enduring struggle of standards-based reform: Lessons from a National Research Center on College and Career-Ready Standards*. Annenberg Institute at Brown University.

Quek, C. L., Wong, A. F. L., Divaharan, S., Liu, W-C., Peer, J., & Williams, M. D. (2007). Secondary school students' perceptions of teacher–student interaction and students' attitudes towards project work. *Learning Environments Research*, *10*, 177–187. https://doi.org/10.1007/s10984-007-9030-3.

Rapoport, A., & Yemini, M. (2020). Citizenship, identity, and education: Re-imagining the Contested Terrain. *Education, Citizenship and Social Justice*, *15*(1), 3–9. https://doi.org/10.1177/1746197919859211.

Rentner, D. S., Scott, C., Kober, N., Chudowsky, N., Chudowsky, V., Joftus, S., & Zabala, D. (2006). *From the capital to the classroom: Year 4 of the No Child Left Behind Act*. Center on Education Policy.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessment: Alternative views of aptitude, achievement, and instruction* (pp. 37–75). Kluwer.

Saavedra, A. R., & Opfer, V. D. (2012). *Teaching and learning 21st century skills: Lessons from the learning sciences*. Asia Society: Partnership for Global Learning and the RAND Corporation. https://asiasociety.org/files/rand-1012report.pdf.

Scheerens, J., Ehren, M., Sleegers, P., & de Leeuw, R. (2012). *OECD Review on evaluation and assessment frameworks for improving school outcomes. Country background report for the Netherlands*. Organisation for Economic Co-operation and Development. https://www.oecd.org/education/school/NLD_CBR_Evaluation_and_Assessment.pdf.

Schwartz, H., Bongard, M., Bogan, E. D., Boyle, A. E., Meyers, D. C., & Jagers, R. J. (2022). *Social and emotional learning in schools nationally and in the Collaborating Districts Initiative: Selected findings from the American Teacher Panel and American School Leader Panel Surveys*. RAND Corporation. https://doi.org/10.7249/RRA1822-1.

Schwartz, S. (2021, June 11). *Map: Where critical race theory is under attack*. *Education Week*. https://www.edweek.org/policy-politics/map-where-critical-race-theory-is-under-attack/2021/06.

Shepard, L. A. (2021). *Ambitious teaching and equitable assessment: A vision for prioritizing learning, not testing*. *American Educator*. https://www.aft.org/ae/fall2021/shepard.

Shepard, L. A., & Dougherty, K. C. (1991). *Effects of high-stakes testing on instruction*. Paper presentation. American Educational Research Association and National Council on Measurement in Education, Chicago.

Stecher, B. (2010). *Performance assessment in an era of standards-based educational accountability*. Stanford Center for Opportunity Policy in Education. https://edpolicy.stanford.edu/sites/default/files/publications/performance-assessment-era-standards-based-educational-accountability_0.pdf.

Stecher, B. M., & Mitchell, K. J. (1995). *Portfolio-driven reform: Vermont teachers' understanding of mathematical problem solving and related changes in classroom practice (CSE Technical Report 400)*. National Center for Research on Evaluation, Standards, and Student Testing. https://cresst.org/wp-content/uploads/TECH400.pdf.

Stecher, B. M., Camm, F., Damberg, C. L., Hamilton, L. S., Mullen, K. J., Nelson, C., Sorenson, P., Wachs, M., Yoh, A., Zellman, G. L., & Leuschner, K. J. (2010). *Toward a culture of consequences: Performance-based accountability systems for public services*. RAND Corporation. https://www.rand.org/pubs/monographs/MG1019.html.

Timberlake, A. (2023, July 12). *Building innovative assessments in an era of accountability*. Center for Assessment. https://www.nciea.org/blog/building-innovative-assessments-in-an-era-of-accountability.
Tyack, D. (1974). *The one best system: A history of American urban education*. Harvard University Press.
U.S. Department of Education. (2009). *Race to the Top program: Executive summary*. https://files.eric.ed.gov/fulltext/ED557422.pdf.
U.S. Department of Education. (2018). *The condition of education 2018*. International Comparisons. National Center for Education Statistics. NCES 2018-144. https://nces.ed.gov/programs/coe/indicator/cns.
U.S. Department of Education. (2022). *Frequently asked questions: Impact of COVID-19 on 2021–2022 accountability systems required under the Elementary and Secondary Education Act of 1965 (ESEA)*. https://oese.ed.gov/files/2022/02/Acct-FAQ-Final-2.11.22-1.pdf.
U.S. Department of Education. (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations*. Office of Educational Technology. https://www2.ed.gov/documents/ai-report/ai-report.pdf.
Vinnakota, R. (2019). *From civic education to a civic learning ecosystem: A landscape analysis and case for collaboration*. Red & Blue Works. https://rbw.civic-learning.org/wp-content/uploads/2019/12/CE_online.pdf.
White House Office of Science and Technology Policy. (2022). *Blueprint for an AI bill of rights: Making automated systems work for the American people*. https://www.whitehouse.gov/ostp/ai-bill-of-rights.
Winthrop, R. (2020, June 4). The need for civic education in 21st-century schools. *Policy2020*. https://www.brookings.edu/policy2020/bigideas/the-need-for-civic-education-in-21st-century-schools.
Woo, A., Lee, S., Prado Tuma, A., Kaufman, J. H., Lawrence, R. A., & Reed, N. (2023). *Walking on eggshells—Teachers' responses to classroom limitations on race- or gender-related topics: Findings from the 2022 American Instructional Resources Survey*. RAND Corporation. https://doi.org/10.7249/RRA134-16.
Yoder, N., Atwell, M. N., Godek, D., Dusenbury, L., Bridgeland, J. M., & Weissberg, R. (2020). *Preparing youth for the workforce of tomorrow: Cultivating the social and emotional skills employers demand*. Collaborative for Academic, Social, and Emotional Learning. https://casel.org/casel-gateway-preparing-workforce-of-tomorrow.
Yuan, K., & Le, V.-N. (2012). *Estimating the percentage of students who were tested on cognitively demanding items through the state achievement tests*. RAND Corporation. https://doi.org/10.7249/wr967.
Zaber, M. A., Karoly, L. A., & Whipkey, K. (2019). *Reimagining the workforce development and employment system for the 21st century and beyond*. RAND Corporation. https://doi.org/10.7249/RR2768.