

# REIMAGINING BALANCED ASSESSMENT SYSTEMS



# REIMAGINING BALANCED ASSESSMENT SYSTEMS

Scott F. Marion, *National Center for the Improvement of Educational Assessment*

James W. Pellegrino, *University of Illinois Chicago*

Amy I. Berman, *National Academy of Education*

Editors

National Academy of Education  
Washington, DC

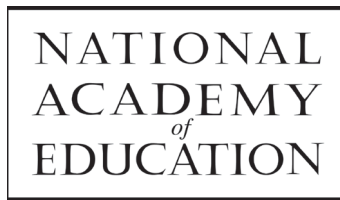
**NATIONAL ACADEMY OF EDUCATION 500 Fifth Street, NW, Washington, DC 20001**

NOTICE: This project and the research reported here were supported by a grant from Smarter Balanced/University of California, Santa Cruz. The opinions expressed are those of the editors and authors and do not represent the views of Smarter Balanced/University of California, Santa Cruz.

Digital Object Identifier: 10.31094/2024/1

Copyright 2024 by the National Academy of Education. All rights reserved.

Suggested citation: Marion, S. F., Pellegrino, J. W., & Berman, A. I. (Eds.). (2024). *Reimagining Balanced Assessment Systems*. National Academy of Education.



The **National Academy of Education** (NAEd) advances high-quality research to improve education policy and practice. Founded in 1965, the NAEd consists of U.S. members and international associates who are elected on the basis of scholarship related to education. The Academy undertakes research studies to address pressing educational issues and administers professional development fellowship programs to enhance the preparation of the next generation of education scholars.

**STUDY ON THE IMPLEMENTATION AND USE OF  
BALANCED ASSESSMENT SYSTEMS**

*Steering Committee*

**Scott F. Marion** (*Co-Chair*), National Center for the Improvement of Educational Assessment

**James W. Pellegrino** (*Co-Chair*), University of Illinois Chicago

**Linda Darling-Hammond**, Learning Policy Institute and Stanford University

**Edward Haertel**, Stanford University

**Jennifer Randall**, University of Michigan

**Lorrie A. Shepard**, University of Colorado Boulder

**Guillermo Solano-Flores**, Stanford University

**James P. Spillane**, Northwestern University

**Jonathan A. Supovitz**, University of Pennsylvania

*Staff*

**Amy I. Berman**, Deputy Director

# Contents

<b>ACKNOWLEDGMENTS</b>	<b>vii</b>
<b>1 REIMAGINING BALANCED ASSESSMENT SYSTEMS: AN INTRODUCTION</b>	<b>1</b>
Scott F. Marion, <i>National Center for the Improvement of Educational Assessment</i> , James W. Pellegrino, <i>University of Illinois Chicago</i> , and Amy I. Berman, <i>National Academy of Education</i>	
<b>2 THE STRUGGLE TO IMPLEMENT BALANCED ASSESSMENT SYSTEMS: EXPLANATIONS AND OPPORTUNITIES</b>	<b>17</b>
Morgan S. Polikoff, <i>University of Southern California</i> , and Ethan L. Hutt, <i>University of North Carolina at Chapel Hill</i>	
<b>3 HUMAN LEARNING AND DEVELOPMENT: THEORETICAL PERSPECTIVES TO INFORM ASSESSMENT SYSTEMS</b>	<b>48</b>
Susan R. Goldman, <i>University of Illinois Chicago</i> , and Carol D. Lee, <i>Northwestern University</i>	
<b>4 CLASSROOM ACTIVITY SYSTEMS TO SUPPORT AMBITIOUS TEACHING AND ASSESSMENT</b>	<b>93</b>
Maria Araceli Ruiz-Primo, <i>Stanford University</i> , and Erin Marie Furtak, <i>University of Colorado Boulder</i>	
<b>5 ASSESSMENT LITERACY AND PROFESSIONAL LEARNING</b>	<b>132</b>
E. Caroline Wylie, <i>National Center for the Improvement of Educational Assessment</i> , and Margaret Heritage, <i>Heritage Consulting</i>	

<b>6</b>	<b>DISTRICT AND SCHOOL PRACTICES AND ASSESSMENTS TO SUPPORT A LEARNING-CENTERED VISION</b>	<b>167</b>
	<i>Elena Diaz-Bilello, University of Colorado Boulder, and Jared Anthony, Carnegie Foundation for the Advancement of Teaching</i>	
<b>7</b>	<b>STATE PRACTICES AND BALANCED ASSESSMENT SYSTEMS</b>	<b>200</b>
	<i>Carla Evans, National Center for the Improvement of Educational Assessment, and Erika Landl, National Center for the Improvement of Educational Assessment</i>	
<b>8</b>	<b>DEVELOPING, IMPLEMENTING, AND INSTITUTIONALIZING COMPLEX EDUCATIONAL INNOVATIONS: CONSIDERATIONS FOR BALANCED ASSESSMENT SYSTEMS</b>	<b>237</b>
	<i>Donald J. Peurach, University of Michigan, and Jennifer Lin Russell, Vanderbilt University</i>	
<b>9</b>	<b>POLICY INFLUENCES ON AMBITIOUS CLASSROOM INSTRUCTION, ASSESSMENT, AND LEARNING</b>	<b>273</b>
	<i>Laura S. Hamilton, American Institutes for Research, and José Felipe Martínez, University of California, Los Angeles</i>	
	<b>BIOGRAPHICAL SKETCHES OF STEERING COMMITTEE MEMBERS, AUTHORS, AND REVIEWERS</b>	<b>309</b>

# Acknowledgments

High-quality assessments are a critical component of the U.S. educational system. Educational assessment can provide information that can be used for making decisions about student learning; teachers, curricula, programs, and schools; resources and funding; and other aspects of educational policy. When used appropriately, educational assessment can serve as an integral tool for the improvement of teaching and learning. No one test, however, can address these multitude of uses and that is why it is critical to have in place *balanced assessment systems*.

This volume explores the history of balanced assessment systems and reimagines balanced assessment systems that center equitable classroom learning environments. In doing so, it provides guidance to state and local educational agencies, as well as schools and teachers, regarding how to (1) foster and maintain a culture of productive assessment use to improve ambitious and equitable teaching and learning at the classroom level; (2) design policy, professional learning, and other local systems necessary to implement balanced assessment systems; and (3) implement processes to use aggregate data to continually improve the assessment system itself to better serve all students, especially those most disenfranchised.

The work leading to this volume was made possible by a generous grant from Smarter Balanced/University of California, Santa Cruz. The National Academy of Education (NAEd) is especially grateful for the support of Tony Alpert, Executive Director; Rochelle Michel, Deputy Executive Program Officer; and Melissa Estuesta, Contract Monitor.

The NAEd steering committee formed to lead this project includes Linda Darling-Hammond, Edward Haertel, Jennifer Randall, Lorrie Shepard, Guillermo Solano-Flores, James Spillane, and Jonathan Supovitz. Each of them committed significant time from their busy schedules to help shape this volume, guide the critical reimagination of balanced assessment systems, and provide numerous rounds of critical feedback to the chapters within this volume.



We want to thank the chapter authors for their critical contributions to this volume. Each chapter stands on its own as an important contribution to the field. It was through the authors' willingness and effort to work across chapters, and with the steering committee and external reviewers, that this collection became a coherent and comprehensive volume.

This volume would not be what it is without the wise and generous support and feedback from our external reviewers:

Courtney Bell, *University of Wisconsin–Madison*  
Kathryn Dewsbury-White, *Michigan Assessment Consortium*  
Debbie Durrence, *Gwinnett County Public Schools*  
John Q. Easton, *Institute for Policy Research at Northwestern University*  
Brian Gong, *National Center for the Improvement of Educational Assessment*  
Ajit Gopalakrishnan, *Connecticut State Department of Education*  
Peter Leonard, *Chicago Public Schools*  
Scott Norton, *Council of Chief State School Officers (retired)*  
William R. Penuel, *University of Colorado Boulder*  
Allison Timberlake, *Georgia Department of Education*

These reviewers gave tremendously of their time and insights. In addition to reviewing specific chapters numerous times, they, along with the steering committee and authors, participated in a two-day workshop to make this volume what it is. For more information about the reviewers, see their biographical sketches at the end of this volume along with the biographical sketches of the authors and steering committee.

This volume and all of its chapters were extensively reviewed and we want to thank Judith Warren Little, Chair of the Standing Review Committee, for overseeing this process.

Finally, we want to acknowledge the work of the NAEd staff: Abigail Bell, Tess Bonnette, Dian Dong, and Gregory White. The staff worked tirelessly from the very conception of the project to bring it to fruition and to coordinate the many moving parts of this process.

We hope this volume contributes to and informs the critical work of classroom and school educators, district and state personnel, federal and state policymakers, and assessment vendors.

Scott F. Marion, *Committee Co-Chair and Volume Co-Editor*  
James W. Pellegrino, *Committee Co-Chair and Volume Co-Editor*  
Amy I. Berman, *Volume Co-Editor*

# Reimagining Balanced Assessment Systems: An Introduction

Scott F. Marion, *National Center for the Improvement of Educational Assessment*

James W. Pellegrino, *University of Illinois Chicago*

Amy I. Berman, *National Academy of Education*<sup>1</sup>

## CONTENTS

INTRODUCTION .....	2
BALANCED ASSESSMENT SYSTEMS, REDEFINED .....	2
THE ORIGINAL INTENTIONS OF BALANCED ASSESSMENT SYSTEMS .....	3
BARRIERS TO IMPLEMENTING BALANCED ASSESSMENT SYSTEMS .....	6
ADVANCES IN UNDERSTANDING HUMAN LEARNING, EQUITY, CULTURE, AND TEACHING.....	7
Human Learning and Development.....	8
Equitable and Culturally Sustaining Dimensions of Assessment.....	8
Ambitious Teaching.....	10
DESIGNING A BALANCED ASSESSMENT SYSTEM.....	11
A ROADMAP FOR THIS VOLUME .....	11
REFERENCES.....	14

---

<sup>1</sup> The authors would like to thank Linda Darling-Hammond and Lorrie Shepard for their very helpful comments and suggestions.

## INTRODUCTION

High-quality assessments are crucial to many aspects of the educational process. They can help policymakers monitor long-term educational trends, assist state educational agencies (SEAs) and local educational agencies (LEAs) in allocating resources and professional development opportunities, provide insights to teachers about how well students have learned the knowledge and skills in an instructional unit, and help teachers and students adjust learning and instruction during daily interactions. Broadly speaking, educational assessment involves tools and processes used to gather information to support a range of decisions—from classroom instruction to school-level professional learning topics to district, state, and federal policies.

Education leaders are regularly bombarded with false claims about assessments that can purportedly serve multiple purposes. These claims feed into misconceptions about the utility of results from certain types of educational tests. However, the harsh reality is that educational assessments are currently designed and validated for a very limited set of purposes and uses—typically only one interpretive use per assessment. The need—to support the full range of uses of assessment information—is the reason why assessment experts and others have called for the design and development of balanced assessment systems, in which the system’s different assessment components complement and support each other.

This volume explores the history of balanced assessment systems and reimagines balanced assessment systems that center equitable classroom learning environments. In doing so, it provides guidance to state and local educational agencies, as well as schools and teachers, regarding how to (1) foster and maintain a culture of productive assessment use to improve ambitious and equitable teaching and learning at the classroom level; (2) design policy, professional learning, and other local systems necessary to implement balanced assessment systems; and (3) implement processes to use aggregate data to continually improve the assessment system itself to better serve all students, especially those most disenfranchised.

## BALANCED ASSESSMENT SYSTEMS, REDEFINED

This volume’s editors, steering committee members, and chapter authors recognize that the definition of balanced assessment systems put forth over two decades ago in *Knowing What Students Know: The Science and Design of Educational Assessment* and by other authors (e.g., Stiggins, 2001) represented an important advance in educational measurement and assessment (National Research Council, 2001). However, interpreting and implementing the vision set forth in *Knowing What Students Know* for both educational assessment and balanced assessment systems has been challenging for many reasons, as is discussed in this chapter and throughout this volume.

Balanced assessment systems and practices, as conceived by this volume’s authors, are intentionally designed to provide feedback to students and information for teachers to support ambitious and equitable instructional and learning opportunities. This type of assessment system facilitates educator engagement in high-leverage professional practices such as quality formative assessment to support ambitious and equitable teaching. Assessments outside of the classroom, at the district and state level, provide aggregate data to policymakers and education leaders, allowing for the monitoring of

educational opportunities and support for high-quality instruction indirectly through the provision of appropriate curricular resources and professional development opportunities. Additionally, these external assessments are designed to coherently support practices that enhance learning and teaching by, among other functions, signaling the types of performance expected in rich and culturally sustaining classroom learning environments. Balanced assessment systems that honor high-quality and equitable classroom learning systems support teachers and educational leaders in improving instructional opportunities and professional practice and may also provide a valuable evidence infrastructure that supports teachers and educational leaders in working within existing systems and interrogating, disrupting, and rebuilding systems to improve instructional opportunities and professional practice.

This volume argues that equitable classroom learning, instruction, and assessment environments must be the focus of balanced assessment systems (see Chapter 4 of this volume, “Classroom Activity Systems to Support Ambitious Teaching and Assessment”). By centering the classroom while developing an assessment system, the components and practices of such systems are more likely to truly support teaching and learning. Therefore, for systems of assessments to be “balanced,” they must support, directly or indirectly, teaching and learning that occurs in the classroom. This assessment system focus is consistent with the current purposes and uses of large-scale assessments—like monitoring long-term educational trends—because, we argue, these district- and state-level assessments provide evidence about program quality, resources, and learning outcomes that can be used to improve those things that affect classroom teaching and learning (see Chapter 6, “District and School Practices and Assessments to Support A Learning-Centered Vision,” and Chapter 7, “State Practices and Balanced Assessment Systems,” of this volume).<sup>2</sup> Once the design and implementation of balanced assessment systems shift to supporting equitable and ambitious classroom learning and instruction, assessment designers must consider, “To what degree and in what ways does this assessment—its content and practices—support or hinder ambitious and equitable classroom learning environments?”

## THE ORIGINAL INTENTIONS OF BALANCED ASSESSMENT SYSTEMS

The call for balanced assessment systems began more than 20 years ago in an effort to correct the distortions and negative effects that occur when large-scale tests are prioritized and often linked to high-stakes decisions. The seminal publication *Knowing What Students Know: The Science and Design of Educational Assessment* included a recommendation that “[t]he balance of mandates and resources should be shifted from an emphasis on external forms of assessment to an increased emphasis on classroom formative assessment designed to assist learning” (National Research Council, 2001, p. 310).

The push for balance signified much more, however, than merely increasing the amount of formal testing done in classrooms to equal the weight of state-level tests. Rather, the intention was to fundamentally change the character of classroom assess-

---

<sup>2</sup> The authors’ definition of balanced assessment systems continues to include coherence, continuity, comprehensiveness, and utility as described in *Knowing What Students Know* and discussed more fully below and in Chapter 2 of this volume, “The Struggle to Implement Balanced Assessment Systems: Explanations and Opportunities,” but emphasizes supporting ambitious and equitable classroom learning and instruction.

ment practices to make them a part of effective teaching and learning. Indeed, the *Knowing What Students Know* study committee was convened to consider how measurement models and assessment methods should be revised in light of current conceptions of learning and knowledge development (National Research Council, 2001). Advances in learning research present in 2001 and even more so today demand fundamental shifts in the representation of authentic learning goals and processes (Nasir et al., 2020; National Academies of Sciences, Engineering, and Medicine, 2018; National Research Council, 2000). At the classroom level, a balanced assessment system will support assessment practices that are thoroughly integrated with day-to-day instructional practices and support deep disciplinary learning. At the level of school districts and states, a balanced assessment system will provide broader aggregate evidence of student attainment to inform policy decisions—including resource allocation.

*Knowing What Students Know* outlined three criteria—coherence, comprehensive-ness, and continuity—to characterize and define balanced assessment systems (National Research Council, 2001). According to *Knowing What Students Know*, systems are balanced when the various assessments are *coherently* linked through a clear specification of the learning targets, *comprehensively* provide multiple sources of evidence to support educational decision-making, and *continuously* document student progress over time (National Research Council, 2001). The study committee believed that these three properties were necessary for creating a high-quality system of assessments rooted in a common model of knowing and learning.

Unfortunately, developers and users have struggled to understand and implement these criteria. In some cases, with a desire to meet the *comprehensiveness* criterion in particular, state and local assessment leaders have overbuilt collections of assessments that can lead to confusion and incoherence. Similarly, to address *continuity*, state and district leaders often turn to quantitative measures of student growth derived from commercial interim or state assessments. Student longitudinal growth measures have value for making comparisons among jurisdictions and over time, and researchers have been working on content-referenced approaches to student growth that focus attention on qualitative distinctions in student learning progress inferred from changes in assessment performance (e.g., Student, 2022). This approach, which is still being investigated, will likely help assessment system developers meet the *continuity* criterion but includes uncommon assessment design requirements.

The *coherence* criterion, which can more readily be understood and operationalized, is, we argue, the most critical of the three criteria for evaluating the quality of balanced assessment systems. The *coherence* criterion signifies the need to connect the various external and classroom assessments with a shared, research-based model of human learning (discussed in greater detail below in the section “Human Learning and Development”). A *coherent* assessment system must be compatible with how student learning is expected to progress within an instructional domain. To work synergistically, assessments at different levels of the educational system must be compatible, although different in grain size or specificity.

An assessment system is *vertically coherent* when there is compatibility among the models of student learning underlying the system’s various assessments (National Research Council, 2006). *Vertical coherence*, based on current conceptions of student learning and anchored in rich classroom learning environments, is a critical consider-



ation for the development of balanced assessment systems. *Knowing What Students Know* promoted the vision of *vertical coherence* among assessments ranging from classroom to state level, but more recent work questions the feasibility of this idea in practice (e.g., Marion, 2018; Shepard et al., 2018). In particular, states' hands-off approach to curriculum and the curriculum-agnostic design of most state assessments makes it difficult for state assessments to coherently connect to a vision of learning and knowing—generally embodied in curriculum documents—for more than a small proportion of school districts in a state. Nevertheless, state assessments, as described in Chapter 7 of this volume, “State Practices and Balanced Assessment Systems,” could, depending on how they are designed, support the *vertical coherence* of district and classroom assessment systems.

At the classroom level, coherence generally means ensuring that assessments are consistent with high-quality curricula and instructional materials that reflect contemporary understandings of disciplinary learning and knowledge development. *Horizontal coherence* is alignment among curriculum, instruction, and assessment to help students develop proficiency in a content domain (National Research Council, 2006). Thus, both *vertical* and *horizontal coherence* are necessary to achieve balanced assessment systems. *Horizontal coherence* is most critical at the classroom level, especially because formative and other classroom assessments must cohere with ambitious instruction and an equity-centered curriculum. School districts generally have the authority to support *horizontally coherent* systems of assessment since curriculum and other related decisions are generally made at the district level.

Many scholars have helped advance the original conceptualization of assessment systems<sup>3</sup> (e.g., Chattergoon & Marion, 2016; Coladarci, 2002; Darling-Hammond et al., 2013; Gong, 2010; National Research Council, 2003, 2006, 2014; Perie et al., 2009; Shepard, 2000; Stiggins, 2001, 2006, 2008). Yet, even with more than 20 years of development and enactment since the publication of *Knowing What Students Know*, there are few examples of well-functioning assessment systems where substantive coherence can be seen among the representations of learning goals at classroom, district, and state levels. There are genuine obstacles that preclude the development of balanced assessment systems, and thus, finding high-quality examples in practice is very rare (Conley, 2018; Marion et al., 2019). The revised definition of balanced assessment systems in this volume is not a major reconceptualization—it is an augmentation because the authors of *Knowing What Students Know*, at the time of the report's publication, could not have anticipated the countervailing forces that arose in response to the changing policy context, including the increasing significance of state-level accountability tests and the proliferation of commercial interim assessment products.

The original call for balanced assessment systems arose from a recognition that most state accountability tests poorly served what should be the primary purpose of assessment: improving learning and instruction. Educators continue to understand that large-scale summative tests are too distal from instruction, at the wrong grain size, and administered at the wrong time of year to make a difference in their daily practice. Nonetheless, following the passage of the No Child Left Behind Act (NCLB) of 2001,

---

<sup>3</sup> For a fuller discussion of this conceptualization, see Chapter 2 of this volume, “The Struggle to Implement Balanced Assessment Systems: Explanations and Opportunities.”

many district leaders turned to commercially available interim assessments marketed to gauge likely results on state-level summative assessments and enhance student achievement. Often, however, these assessments do not clearly link to other levels of the assessment system and the results do not help improve student learning (Perie et al., 2009). Therefore, the renewed call for balanced assessment systems made by this volume is motivated by the desire to enhance the utility of assessments for improving learning and instruction as well as for monitoring, accountability, and evaluation purposes.

## BARRIERS TO IMPLEMENTING BALANCED ASSESSMENT SYSTEMS

Why don't more balanced assessment systems exist in practice if there is such a need? We describe some of the key hurdles that have made it challenging to enact high-quality balanced assessment systems to help leaders recognize and overcome these barriers as they engage in the design or redesign of assessment systems (see Table 1-1).

**TABLE 1-1** Key Barriers to Implementation of Balanced Assessment Systems<sup>a</sup>

- 
1. Influence of politics, policy, and political boundaries
  2. Influence of commercialization and proliferation of assessments
  3. Lack of attention to curriculum and learning in the design of assessment systems
  4. Lack of assessment literacy at multiple levels of the system
  5. Limited understanding of human development and student learning
  6. Misconceptions associated with the meaning of balance
- 

<sup>a</sup> This table represents critical barriers to the implementation of balanced assessment systems. The first four points are discussed more fully in Marion et al. (2019). The final two are further explored in this chapter. Additionally, in Chapter 2 of this volume, "The Struggle to Implement Balanced Assessment Systems: Explanations and Opportunities," Polikoff and Hutt discuss and expand on these and other barriers, which they categorize as political or technical.

Turning first to the influence of politics and policy on balanced assessment systems, NCLB, which mandated high-stakes testing for all U.S. schools, was enacted only a few months after the publication of *Knowing What Students Know*. Consequently, most states rushed to design and implement a set of state-wide assessments, creating annual accountability tests in grades 3–8 and at least once in high school to comply with NCLB. Furthermore, leadership at the U.S. Department of Education encouraged states to save money and time, especially given the amount of new testing required, by relying almost exclusively on multiple-choice items to populate their end-of-year tests. This shift away from a variety of constructed-response and performance-based assessments to an almost exclusive use of multiple-choice tests increased the incoherence of state assessment systems because the content of such tests was poorly aligned with what was known about how students should learn critical aspects of disciplinary knowledge and skill. These efforts were, in large part, incompatible with the notions of balanced assessment systems put forth in *Knowing What Students Know*.

In addition to NCLB and the associated onslaught of federally mandated testing, Marion and colleagues (2019) describe in detail the influence of the commercialization and proliferation of assessments, lack of attention to curriculum and learning in the design of assessment systems, and lack of assessment literacy concerning how to implement and use assessment information. Moreover, two other critical barriers to balanced assessment systems have shaped this volume: limited understanding of human development and student learning, and misconceptions associated with the meaning of balance.

A limited understanding of human development and student learning has led to a lack of coherence between the design of assessments generally—as well as systems of such assessments—and the knowledge and skills that tests should be assessing. While *Knowing What Students Know* called for balanced assessment systems to be coherently connected via a common model of knowing and learning, this call did not mean that any model of learning would be acceptable. Rather, following the publication of *How People Learn: Brain, Mind, Experience, and School* in 2000, *Knowing What Students Know*, in 2001, envisioned contemporary, research-based theories of learning and cognition anchoring the coherence of the design and development of balanced systems of assessment (National Research Council, 2000). Achieving coherence with modern conceptions of knowing and learning requires information at different grain sizes to support the development of deep disciplinary knowledge or to monitor long-term educational trends. Unfortunately, far too few assessment designers and educators deeply understand the process of how students come to develop knowledge and skills within and across subject areas, which is why this volume is grounded in an explication of the sciences of human development and learning that have accrued since the publication of *How People Learn* (see Chapter 3 of this volume, “Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems”).

Finally, balanced assessment systems have been constrained in practice due to varied interpretations of the meaning of balance. Even though the original vision of *Knowing What Students Know* called for coherence from the schoolhouse to the state house, the NCLB-initiated accountability pressures associated with the use of state tests resulted in an outsized emphasis on state assessments. This led to “teaching to the test” for large-scale state standardized tests as well as significant time spent on test preparation and testing “tricks,” particularly in historically marginalized communities, instead of focusing on curriculum-rich classroom teaching and learning supported by formative assessment practices (e.g., Shepard et al., in press). Adding to this imbalance, many assessment companies promoted the notion that commercial interim assessments are an essential component of any assessment system, further tilting the concept of balance away from classroom assessment and learning.

A common image of assessment systems—often represented as a seesaw with state assessments at one end, classroom assessments at the other, and interim assessments at the fulcrum—has had negative repercussions in terms of developing high-quality classroom assessments and instantiating formative assessment practices. The time, energy, and money devoted to state-wide and commercial interim assessments, along with their perceived value relative to classroom assessments, have detracted from efforts to develop high-quality classroom assessment resources and professional learning programs to support the development of formative assessment literacy among educators. This volume offers a different image—one that is centered on ambitious and equitable classroom learning environments supported by balanced assessment practices.

## **ADVANCES IN UNDERSTANDING HUMAN LEARNING, EQUITY, CULTURE, AND TEACHING**

The years since 2000 have seen many changes in widely held conceptions of assessment, equity, student learning, and instruction that must be incorporated into the new



vision of balanced assessment systems. Three critical advances include further conceptualizations of (1) human learning and development, (2) equitable and culturally sustaining dimensions of assessment, and (3) ambitious teaching.

### **Human Learning and Development**

Since the publication of *Knowing What Students Know* and *How People Learn: Brain, Mind, Experience, and School*, there have been numerous advances in the science of human learning and development. *How People Learn II: Learners, Context, and Cultures* summarized emerging theory and research emphasizing the social nature of human learning and the importance of cultural and linguistic backgrounds in shaping what individuals know and how they learn (National Academies of Sciences, Engineering, and Medicine, 2018). *How People Learn II* also described how the fields of cognitive and developmental neuroscience have provided considerable insights into how learners develop competence in given domains. These advances in theory and research on the nature of human development and learning and how this new knowledge relates to assessment and assessment systems are the foci of Chapter 3 of this volume, “Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems.” Moreover, key aspects of research on human learning and development are attended to in this volume’s other chapters as authors discuss assessments designed to directly support student learning and ambitious teaching (Chapter 4, “Classroom Activity Systems to Support Ambitious Teaching and Assessment”) or assessments intended to support the needs of education leaders and policymakers (Chapter 6, “District and School Practices and Assessments to Support a Learning-Centered Vision,” and Chapter 7, “State Practices and Balanced Assessment Systems”).

### **Equitable and Culturally Sustaining Dimensions of Assessment**

The years since 2000 have also seen a greater urgency in understanding how assessment can support or hinder equity and social justice. *Knowing What Students Know* did highlight equity goals: “Issues of fairness and equity must be central concerns in any effort to develop new forms of assessment. Relevant to these issues is a substantial body of research on the social and cultural dimensions of cognition and learning” (National Research Council, 2001, p. 32). However, there has been expansive scholarship since then that has elaborated on topics such as culturally relevant and sustaining pedagogy, social justice, and equity, which enriches understanding of education in general and assessment specifically.

Most people working in education agree that “educational equity” is an important aim of schooling. However, the almost universal acknowledgment that equity is a valuable goal can obscure very real differences in what various people and organizations mean by “equity” and how they operationalize it. (Levinson et al., 2022, p. 1)

Equity can focus on resources, opportunities, and/or outcomes. In terms of resource allocation and opportunity to learn, the authors of this volume use the National Academies of Sciences, Engineering, and Medicine’s (2019) definition of equity as put forth in the report *Monitoring Educational Equity* as a foundation:

Educational equity requires that educational opportunity be calibrated to need, which may include additional and tailored resources and supports to create conditions of true educational opportunity.... This idea of equity is different from equality, which connotes the idea that certain goods and services are distributed evenly, irrespective of individual needs or assets.

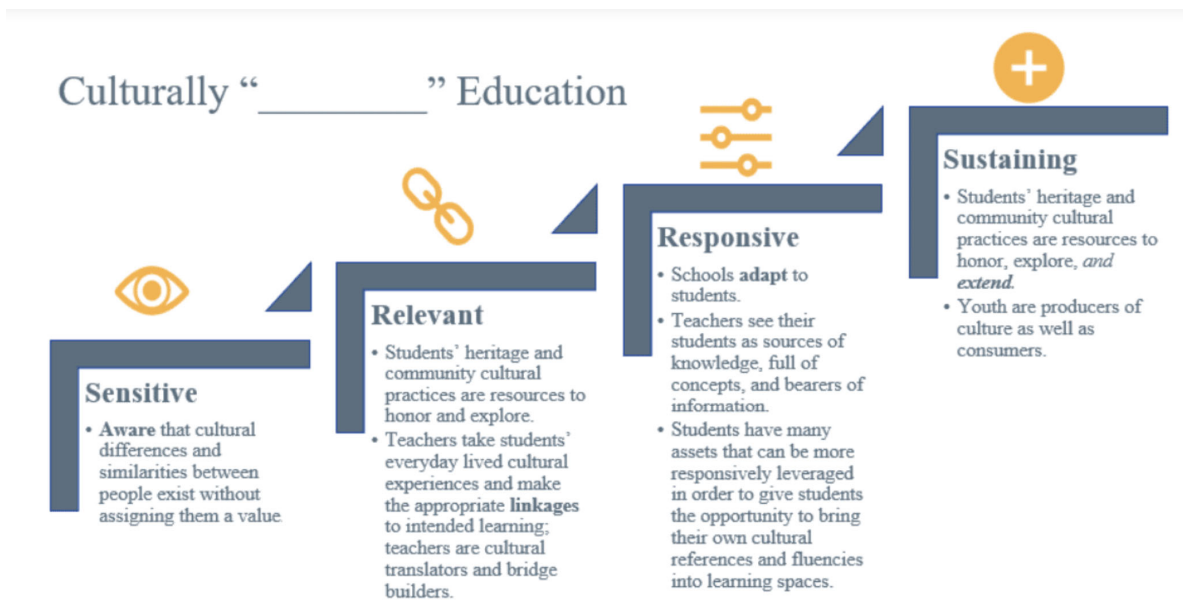
The circumstances in which students live affect their academic engagement, academic progress, and educational attainment in important ways. If narrowing disparities in student outcomes is an imperative, schools cannot shirk the challenges arising from context.... For education, [this requires the] meaningful examination of equity between key population groups, such as those defined by socioeconomic status, race and ethnicity, or English proficiency, ... [and includes an examination of measures of] disparities in students' academic achievement and attainment outcomes and engagement in schooling ... [as well as] access to resources and opportunities.... (National Academies of Sciences, Engineering, and Medicine, 2019, pp. 2–3)

This definition focuses on resources and outcomes, which are undeniably important, but educators must also embrace students' linguistic and cultural heritages as essential aspects of effective instruction and assessment. While there is a long history of addressing inequities in pedagogy (e.g., Ladson-Billings, 1995; Moll et al., 1992) and assessment (e.g., Gordon, 1995), there has been a recent growing wave of recognition of the need to embrace and incorporate students' linguistic, cultural, ethnic, and racial backgrounds in curriculum, instruction, and assessment to support all students' social and intellectual development more equitably (e.g., Paris, 2012; Randall et al., 2022). This volume was written from the perspective that balanced assessment systems that support rich classroom learning contexts must be designed to explicitly support equity and social justice.

Shifting to more equitable assessment practices and balanced assessment systems also requires shifting assessment design and implementation to approaches that reflect greater cultural awareness. Building from the seminal work of Gloria Ladson-Billings (1995), Geneva Gay (2002), Django Paris (2012), and others, Carla Evans (2021) summarized the various current terms related to the concept of cultural awareness contributing to the effort to make assessments more equitable and just—culturally sensitive, culturally relevant, culturally responsive, and culturally sustaining (see Figure 1-1, a stair-step illustration of these terms).

All four concepts emphasize incorporating the cultural and linguistic knowledge and practices that students bring to school as a means of making instruction and curriculum more engaging. In classrooms that embrace these approaches, “teachers explore their students' cultural and social identities and make connections with students' communities; they get to know individual students, their families, and the values, beliefs, practices, and funds of knowledge each student brings to the classroom” (Taylor & Nolen, 2022, p. 58). Furthermore, valuing and incorporating students' cultural and linguistic heritages in instruction allows students to question existing power structures and envision a different social order (Paris, 2012).

These concepts and their implications for curriculum, instruction, and assessment are explored and expanded throughout this volume—especially in Chapter 3, “Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems;” Chapter 4, “Classroom Activity Systems to Support Ambitious Teaching;” and Chapter



**FIGURE 1-1** Culturally “\_\_\_\_\_” Education.  
SOURCE: Evans (2021).

5, “Assessment Literacy and Professional Learning.” For example, Wylie and Heritage write in Chapter 5 of this volume, “Assessment Literacy and Professional Learning”: “Achieving equity requires a culturally sustaining approach to pedagogy and a fair and just approach to assessment, including interrogating the content of what is taught and how it is taught, together with what and how that content is assessed” (p. 133). The notion of “interrogating,” like social justice, is action-oriented.

### **Ambitious Teaching**

Consistent with advances in theory and research on human development and learning, as well as a focus on equity and culture, this volume envisions assessments as supporting ambitious teaching in classrooms. Ambitious teaching, grounded in socio-cultural theory, calls for deeply knowing the multiple dimensions of each student—academic, emotional, social, and cultural—and providing a supportive and nurturing classroom environment where students feel safe to talk together about their thinking, reasoning, and identities within disciplinary communities of knowledge and practice. Ambitious teaching intentionally aims to empower all students to use the disciplinary knowledge and skills they acquire to solve authentic problems (Ball & Forzani, 2009; Lampert & Graziani, 2009; Shepard, 2021). Equitable assessments that reveal the depths of students’ thinking are a critical component of an ambitious teaching environment and are used to support each student’s learning and development while providing valuable instructional insights for educators (see Chapter 4 of this volume, “Classroom Activity Systems to Support Ambitious Teaching,” for additional discussion on ambitious teaching).

## DESIGNING A BALANCED ASSESSMENT SYSTEM

Designing assessments to cohere as part of a system requires intentional and thoughtful approaches to ensure assessments can support—or at least not hinder—classroom instruction and assessment. A theory of action is a useful heuristic to support this type of design work because it provides a comprehensive framework for analysis, evaluation, and continuous improvement. A theory of action can explain and then guide the interactions among the components of the assessment system to maximize opportunities for the various assessments to support the system’s common vision of learning and goals (Bennett, 2010; Chattergoon & Marion, 2016).

A theory of action describes the inputs, processes, mechanisms, and intermediate steps necessary to realize the goals. In other words, it is not enough to announce that an assessment system will improve learning and teaching. Rather, developers must understand—and clearly communicate—how the proposed assessment, or set of assessments, will support desired changes in teaching and learning. Developers should ask themselves: what activities and resources need to be put in place to maximize the chances of realizing the intended outcomes?

This is challenging design work for single assessments and is that much more complex when trying to design entire balanced assessment systems. System designers need to rely on a well-specified theory of action to ensure that the various components of the system meet the needs and uses of various stakeholders. Such a theory of action should be created in a way that prompts designers to reflect upon the criteria for balanced assessment systems discussed above.

## A ROADMAP FOR THIS VOLUME

Drawing on the framework and reimagining of balanced assessment systems outlined in this chapter, this volume aims to provide a roadmap for developing, implementing, and using balanced assessment systems to support ambitious and equitable teaching and learning. The volume documents prior struggles in implementing balanced assessment systems (Chapter 2); expounds the theoretical underpinnings of human learning and development (Chapter 3); and situates the work of balanced assessment systems within classrooms supporting ambitious and equitable teaching and learning with robust assessment literacy and professional learning for educators (Chapters 4 and 5). At the same time, it recognizes the critical roles of schools, districts, and states in establishing and supporting balanced assessment systems (Chapters 6 and 7). It also provides considerations for developing, implementing, and institutionalizing the complex educational innovation of balanced assessment systems, as well as critical lessons for enacting policies to promote balanced assessment systems (Chapters 8 and 9). While the chapters are individually authored, the steering committee, chapter authors, and additional chapter reviewers (including representatives from SEAs and LEAs) spent significant time working together to outline the volume, review the chapters, and ensure that through these chapters, the entire volume provides a roadmap to developing balanced assessment systems centered on ambitious and equitable teaching and learning.

Chapter 2, “The Struggle to Implement Balanced Assessment Systems: Explanations and Opportunities,” provides critical background information on the origins of bal-

anced assessment systems and barriers to their implementation. The chapter discusses the historical context in which balanced assessment systems emerged, as well as the original principles of such systems and how they evolved and were operationalized over time. While the chapter details the technical and political/practical challenges that have hindered the implementation of balanced assessment systems, it also reviews several efforts to implement elements of these systems and highlights lessons that can be gleaned from the examples.

Chapter 3, “Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems,” articulates why balanced assessment systems will yield the most useful information if they are informed by a comprehensive understanding of the complexities of human learning and development. The chapter provides the foundational principles of human learning and development and their implications for supporting robust, anti-racist learning environments through teaching and assessment practices. The chapter provides case studies to demonstrate how teaching, learning, and assessment connect to students’ knowledge and repertoires through their participation in everyday routine cultural practices. Understanding the multiple pathways through which humans, as individuals and communities, engage in sense-making, problem-solving, and learning is critical to determining assessment validity.

Chapter 4, “Classroom Activity Systems to Support Ambitious Teaching and Assessment,” builds on sociocultural theories of learning to conceptualize a learning environment that supports ambitious classroom teaching and assessments. Attending to equity and culture, the chapter describes and provides rich examples of the elements comprising a classroom activity system that supports ambitious teaching and assessments—learners (including their interests, identities, linguistic and cultural capital, and knowledge about themselves), curriculum, instruction, learning culture, and assessment. The chapter explores assessments as a process of gathering, analyzing, and interpreting relevant information about where students are in relation to rich learning goals and provides examples of formative and summative assessment practices that support said goals and embody a deep understanding of student learning, levels of knowledge, skills, and practices. The chapter also provides design features of classroom assessment to support ambitious instruction, including attention to cultural and social relevance, fairness and representation, and cognitive demands.

Chapter 5, “Assessment Literacy and Professional Learning,” examines what assessment literacy means within the reconceptualization of assessment practices outlined in this volume as well as how it can be promoted among and engaged by teachers. The chapter addresses how assessment literacy can facilitate equitable and just learning outcomes. It also identifies the knowledge and skills teachers need to make effective use of classroom assessments, including developing learning goals, generating assessment evidence, interpreting student responses, and guiding decisions intended to advance the learning and development of each student. The chapter then describes the three enabling conditions that ground teachers’ professional learning for developing assessment literacy competencies—sociocultural consciousness and agency, learning supports, and deliberate practice. The chapter operationalizes these enabling conditions in teachers’ local settings, supported by strong, collaborative peer learning communities. Similarly, the chapter outlines how school and district leaders and state policy play pivotal roles in providing systemic support for assessment literacy.



Chapter 6, “District and School Practices and Assessments to Support a Learning-Centered Vision,” discusses the practices and strategies that schools and districts can utilize to support and sustain assessments focused on ambitious teaching and learning. This chapter briefly describes the traditional roles school districts play in influencing instructional work in individual schools and then posits what it would look like for a *learning system district* to support assessments used to support ambitious teaching and learning (see Chapter 8 of this volume, “Developing, Implementing, and Institutionalizing Complex Educational Innovations: Considerations for Balanced Assessment Systems,” for additional information about *learning system districts*). In this situation, schools and districts would prioritize the use of classroom assessments, and the chapter expounds on the strong, supportive infrastructure this would require—including high-quality curricula, professional learning, and grading policies. The chapter also addresses how districts might begin working with schools to move toward this vision for teaching and learning while also engaging in necessary evaluations to monitor the implementation of this work.

Chapter 7, “State Practices and Balanced Assessment Systems,” acknowledges that states cannot design or implement balanced assessment systems on their own and argues that the primary role of states in promoting such systems is to support the right structures and conditions for districts, schools, and classroom educators to do their jobs effectively and improve student learning. The chapter situates the state’s role within the larger sociopolitical context—specifically how federal accountability and peer review requirements influence state assessment decisions and exert pressure on districts, schools, and ultimately classrooms. The chapter articulates several state actions that contribute to “imbalance” and defines what is under local versus state control regarding the design and implementation of balanced assessment systems. The chapter then provides six high-leverage actions that SEAs can take to support local efforts to design and implement balanced assessment systems: (1) set a clear, compelling, and coherent theory of action for balanced assessment systems, (2) clearly communicate the intended role of the state summative assessment and other state-provided resources, (3) proactively design state content standards, curriculum frameworks, and state assessments to promote coherence, (4) mitigate misuse of state tests through clear reporting and guidance, (5) provide tools, resources, and supports to LEAs, and (6) engage educator preparation programs.

Chapter 8, “Developing, Implementing, and Institutionalizing Complex Educational Innovations: Considerations for Balanced Assessment Systems,” utilizes a larger analytical framework to contextualize the potential challenges of implementing balanced assessment systems and then models the use of this framework for understanding and addressing the complexities of such systems. Chapter 8 builds on the earlier chapters in this volume to situate the definition, goals, and challenges of balanced assessment systems in a historical context. The chapter presents a historical analysis of the accumulation and co-evolution of policy logics, presses, and local capabilities seeking to advance ambitions for educational quality and equity that are central to balanced assessment systems. The chapter examines accumulating policy logics at the national level that have focused on resources, practice, and empowerment as key levers for advancing educational quality and equity. It also examines how these policy logics have, in turn, pressed local districts to maintain their structural/procedural,

technical, and moral legitimacy. The chapter then examines how these national-level policies and presses have driven the local-level evolution of districts as *school* systems, *education* systems, and *learning* systems, characterized by different functional capabilities for organizing, managing, and improving instruction to advance educational quality and equity. The chapter argues that this analytical framework can be viewed as a developmental sequence useful for analyzing the progress of individual states and local districts in implementing balanced assessment systems. The chapter argues that how balanced assessment systems will function depends on how state and local leaders engage in collaborative learning to craft coherent visions for developing, implementing, and institutionalizing balanced assessment systems based on their current capabilities and contexts.

Chapter 9, “Policy Influences on Ambitious Classroom Instruction, Assessment, and Learning,” builds on and updates the research concerning policy influences on teaching, learning, and assessment, both in the context of balanced assessment systems and more generally. The chapter provides a brief history of policies related to assessment and explores the limitations of previously enacted policies to promote ambitious teaching and high-quality and equitable learning opportunities for all students. The chapter then 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, including (1) adopting an inclusive, collaborative approach to policy design and implementation, (2) interrogating the values that underlie policy, (3) ensuring that state policies are informed by an understanding of local variation, (4) reducing the state assessment footprint and prioritizing coherence and measures that will inform improvement, (5) embracing technological innovation cautiously and responsibly, and (6) recognizing the limits and risks of assessment policy and providing supports for navigating politics.

## REFERENCES

- Ball, D. L., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education*, 60(5), 497–511. <https://doi.org/10.1177/0022487109348479>.
- Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research and Perspectives*, 8(2–3), 70–91. <https://doi.org/10.1080/15366367.2010.508686>.
- Chattergoon, R., & Marion, S. (2016). *Not as easy as it sounds: Designing a balanced assessment system*. National Association of State Boards of Education. <https://csaa.wested.org/wp-content/uploads/2019/11/Chattergoon-Marion.pdf>.
- Coladarci, T. (2002). Is it a house ... or a pile of bricks? Important features of a local assessment system. *Phi Delta Kappan*, 83(10), 772–774.
- Conley, D. T. (2018). *The promise and practice of next generation assessment*. Harvard Education Press.
- Darling-Hammond, L., Herman, J., Pellegrino, J., Abedi, J., Abner, J. L., Baker, E., Bennett, R., Gordon, E., Haertel, E., Hakuta, K., Ho, A., Linn, R. L., Pearson, P. D., Popham, J., Resnick, L., Schoenfeld, A. H., Shavelson, R., Shepard, L. A., Shulman, L., & Steele, C. M. (2013). *Criteria for high-quality assessment*. Stanford Center for Opportunity Policy in Education, Stanford University; Center for Research on Student Standards and Testing, University of California at Los Angeles; and Learning Sciences Research Institute, University of Illinois at Chicago. [https://edpolicy.stanford.edu/sites/default/files/publications/criteria-higher-quality-assessment\\_2.pdf](https://edpolicy.stanford.edu/sites/default/files/publications/criteria-higher-quality-assessment_2.pdf).

- Evans, C. (2021). *Culturally sensitive, relevant, responsive, and sustaining assessment: Are there limits to making large-scale standardized testing culturally responsive?* Center for Assessment. <https://www.nciea.org/blog/culturally-sensitive-relevant-responsive-and-sustaining-assessment>.
- Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education*, 53(2), 106–116. <https://doi.org/10.1177/0022487102053002003>.
- Gong, B. (2010). *Using balanced assessment systems to improve student learning and school capacity: An introduction*. Council of Chief State School Officers and Renaissance Learning.
- Gordon, E. W. (1995). Toward an equitable system of educational assessment. *The Journal of Negro Education*, 64(3), 360–372. <https://doi.org/10.2307/2967215>.
- Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3), 465–491. <https://doi.org/10.2307/1163320>.
- Lampert, M., & Graziani, F. (2009). Instructional activities as a tool for teachers' and teacher educators' learning. *The Elementary School Journal*, 109(5), 491–509. <https://doi.org/10.1086/596998>.
- Levinson, M., Geron, T., & Brighthouse, H. (2022). Conceptions of educational equity. *AERA Open*, 8. <https://doi.org/10.1177/23328584221121344>.
- Marion, S. F. (2018). The opportunities and challenges of a systems approach to assessment. *Educational Measurement: Issues and Practice*, 37(1), 45–48. <https://doi.org/10.1111/emip.12193>.
- Marion, S., Thompson, J., Evans, C., Martineau, J., & Dadey, N. (2019, April 4–8). *A tricky balance: The challenges and opportunities of balanced systems of assessment*. Paper presentation. National Council on Measurement in Education Annual Meeting, Toronto, Ontario, Canada. <https://www.nciea.org/library/tricky-balance-challenges-and-opportunities-balanced-systems-assessment>.
- Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory into Practice*, 31(2), 132–141.
- Nasir, N. S., Lee, C. D., Pea, R., & McKinney de Royston, M. (Eds.). (2020). *Handbook of the cultural foundations of learning*. Routledge: Taylor & Francis Group.
- National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. The National Academies Press. <https://doi.org/10.17226/24783>.
- National Academies of Sciences, Engineering, and Medicine. (2019). *Monitoring educational equity*. The National Academies Press. <https://doi.org/10.17226/25389>.
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academy Press. <https://doi.org/10.17226/9853>.
- 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. (2003). *Assessment in support of instruction and learning: Bridging the gap between large-scale and classroom assessment: Workshop report*. The National Academies Press. <https://doi.org/10.17226/10802>.
- National Research Council. (2006). *Systems for state science assessment*. The National Academies Press. <https://doi.org/10.17226/11312>.
- National Research Council. (2014). *Developing assessments for the Next Generation Science Standards*. The National Academies Press. <https://doi.org/10.17226/18409>.
- No Child Left Behind Act of 2001, 20 USC § 6301. (2001). <https://www.congress.gov/bill/107th-congress/house-bill/1/text>.
- Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher*, 41(3), 93–97. <https://doi.org/10.3102/0013189X12441244>.
- Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, 28(3), 5–13. <https://doi.org/10.1111/j.1745-3992.2009.00149.x>.
- Randall, J., Slomp, D., Poe, M., & Oliveri, M. E. (2022). Disrupting white supremacy in assessment: Toward a justice-oriented, antiracist validity framework. *Educational Assessment*, 27(2), 170–178. <https://doi.org/10.1080/10627197.2022.2042682>.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. <https://doi.org/10.3102/0013189X029007004>.
- Shepard, L. A. (2021). Ambitious teaching and equitable assessment: A vision for prioritizing learning, not testing. *American Educator*, Fall. <https://www.aft.org/ae/fall2021/shepard>.



- Shepard, L. S., Marion, S. F., & Saldaña, C. M. (In press). Standards-based reform and school accountability. In L. Cohen-Vogel, J. Scott, & P. Youngs (Eds.), *Handbook of education policy research* (Vol. 2). American Educational Research Association.
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, 37(1), 21–34. <https://doi.org/10.1111/emip.12189>.
- Stembridge, A. (2020). *Culturally responsive education in the classroom: An equity framework for pedagogy*. Routledge, Taylor & Francis Group.
- Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5–15. <https://doi.org/10.1111/j.1745-3992.2001.tb00065.x>.
- Stiggins, R. (2006). *Balanced assessment systems: Redefining excellence in assessment*. Educational Testing Service. <https://famemichigan.org/wp-content/uploads/2018/06/Stiggins-Balanced-Assessment-Systems-Redefining-Excellence-in-Assessment.pdf>.
- Stiggins, R. (2008). *Assessment manifesto: A call for the development of balanced assessment systems*. Educational Testing Service. <https://famemichigan.org/wp-content/uploads/2018/06/Stiggins-Assessment-Manifesto-A-Call-for-the-Development-of-Balanced-Assessment-Systems.pdf>.
- Student, S. R. (2022, April 22–25). *Calibrating and validating the uses of a vertical scale in a computerized adaptive setting*. Presentation. National Council on Measurement in Education Annual Meeting, San Diego, CA.
- Taylor, C. S., & Nolen, S. B. (2022). *Culturally and socially responsible assessment: Theory, research, and practice*. Teachers College Press.

# The Struggle to Implement Balanced Assessment Systems: Explanations and Opportunities

Morgan S. Polikoff, *University of Southern California*  
Ethan L. Hutt, *University of North Carolina at Chapel Hill*

## CONTENTS

INTRODUCTION .....	18
CONTEXT FOR AND ORIGINS OF THE CONCEPT OF BALANCED ASSESSMENT SYSTEMS .....	19
WHAT ARE BALANCED ASSESSMENT SYSTEMS? .....	21
The Principles of Balanced Assessment Systems.....	23
Updates to the Balanced Assessment Criteria.....	27
THE IMPLEMENTATION OF BALANCED ASSESSMENT SYSTEMS OVER TIME .....	29
New Hampshire’s Performance Assessment of Competency Education.....	29
Learning Progressions.....	32
Smarter Balanced Assessment Systems.....	33
Other Examples in Practice .....	33
FACTORS THAT HAVE HINDERED THE GROWTH AND IMPACT OF BALANCED ASSESSMENT SYSTEMS .....	34
Technical Challenges .....	35
Political and Practical Challenges .....	37
CONCLUSION .....	43
REFERENCES .....	44

## INTRODUCTION

Balanced assessment systems have not yet been as broadly implemented as their proponents desired. Assessment systems throughout the United States are still broadly characterized by incoherence, limited instructional utility, and at best, a modest impact on student learning. The failure of balanced assessment systems to gain a foothold in the standardized testing system that has otherwise massively expanded since the 2001 publication of the National Research Council's (NRC's) *Knowing What Students Know* raises important questions: Why have the ideas behind balanced assessment systems failed to achieve substantial impact in practice? What could be done differently if the goal is achieving greater implementation and impact of balanced assessment systems?

Our answer is that the history of balanced assessment systems underscores an important lesson in school reform: technical superiority is never sufficient to ensure adoption or implementation. Despite the backing of prominent experts and organizations as well as providing a sophisticated approach to assessment, balanced assessment systems have been implemented in only a handful of places and with only limited fidelity to the vision laid out in *Knowing What Students Know*. These developments point to the importance of thinking not just about the technical merit of the *Knowing What Students Know* framework but also about the political and organizational support necessary to secure ongoing implementation. The radical transformation of a country's educational assessment system would be a difficult task under any circumstances, but is considerably more difficult in the United States due to its decentralized structures and byzantine governance that determine assessment policies. Detailing and addressing the challenges posed by the political and organizational realities of American schooling is a crucial step in identifying a possible path for the implementation of balanced assessment systems.

In this chapter, we first briefly discuss the historical context in which the idea of balanced assessment systems emerged. Next, we describe the major tenets of balanced assessment systems as originally conceived and then consider how they have evolved and been operationalized over time. Following this, we describe attempts to implement balanced assessment systems in the past 20 years, and we consider how well they have been implemented and to what effect. Finally, we offer a set of high-level explanations for why balanced assessment systems have failed to take hold, 20 years after the initial ideas were put forth.

We note at the outset that the review and appraisal of balanced assessment systems offered below is hampered by a lack of clarity and common terminology in the field. Even the authors of this volume grappled with this issue as we were collectively writing the book, devoting time to debating key terms and definitions and even whether to use the term "balanced" at all. There are also several related terms that are used at least as frequently as "balanced assessment systems," including "comprehensive assessment systems" (e.g., Brookhart, 2013) and "next-generation assessment" (Conley, 2018). These terms map onto similar—but not identical—intellectual terrain, further obscuring the meaning of any of the three terms. A lack of clarity about the extent of this overlap, as well as the terms' relationship to each other (i.e., what, if anything, is signified in using one term rather than another) has contributed to a sense that the field has failed to cohere over time.

Compounding the challenge of inconsistent terminology, which makes identifying relevant research difficult, much of the work that has been done in this area—especially work at the district level—has not been published. The lack of published research has relegated much of the most important on the ground experience with attempting implementation of balanced assessment systems to the realm of anecdotes and secondhand knowledge. Even putting aside the issue of accuracy or general applicability of these accounts, since they are unpublished, they remain largely inaccessible to interested researchers or practitioners. Thus, this review relies, at least in part, on the research that the assessment experts leading the creation this volume suggested we include, even if that research does not claim to be about balanced assessment systems per se.

### CONTEXT FOR AND ORIGINS OF THE CONCEPT OF BALANCED ASSESSMENT SYSTEMS

At first glance, it seems fitting that a report like *Knowing What Students Know* would be published the same year that Congress enacted a law intended to provide the public with more, and more precise, information about “what students know” than ever before (No Child Left Behind Act, 2001). Although the ideas behind *Knowing What Students Know* and the No Child Left Behind Act of 2001 (NCLB) were developed concurrently, the two documents were created by different groups and reflect very different visions of the future of school assessment in America. The political actors supporting NCLB saw an opportunity to use the federal government’s standardizing power to combat “the soft bigotry of low expectations” (Bush, 2000) through accountability driven by clear standards, annual statewide assessments, and explicit reporting and progress requirements; while the scholars behind *Knowing What Students Know* drew on research on learning and recommended a shift in “the balance of mandates ... from an emphasis on external forms of assessment to an increased emphasis on classroom formative assessment” (National Research Council, 2001, p. 14). At the very moment scholars sought to make assessments that were more authentic and proximate to everyday school practices, envisioning “that assessments at all levels—from classroom to state—will work together in a system that is comprehensive, coherent, and continuous” (National Research Council, 2001, p. 9), federal policy was pulling toward more remote assessments that were more aligned to statewide standards. As the intervening two decades have made clear, the ideas from *Knowing What Students Know* have remained adrift in a political environment focused on external accountability.

Although NCLB and *Knowing What Students Know* embodied different views of assessment, both were attempts to address long-running concerns about the achievement of American students and the capacity of American schools to meet the challenges of a changing world. Indeed, *Knowing What Students Know* articulates two core concerns facing American schools at that time. First, as *Knowing What Students Know* expounds on, there was a view that what mattered in terms of educational learning had shifted profoundly during the two decades prior to its publication. Following the economic malaise of the 1970s, the American economy was transitioning from manufacturing to service jobs. Although American manufacturing represented more than one-fifth of nonfarm jobs in 1979, these jobs would never again represent such a large portion of employment (Harris, 2020). The shift from manufacturing to service was understood

to require a considerable change in what school curricula valued and assessed. The upshot, as *Knowing What Students Know* states, was that,

To succeed in this increasingly competitive economy, all students, not just a few, must learn how to communicate, to think and reason effectively, to solve complex problems, to work with multidimensional data and sophisticated representations, to make judgments about the accuracy of masses of information, to collaborate in diverse teams, and to demonstrate self-motivation. (National Research Council, 2001, p. 22)

These changes, all associated with the rise of the information economy, were also supposed to be amplified by technological shifts involving the increased capacities of computers, the Internet, and electronic communication like email.

Second, *Knowing What Students Know* argued that American students needed to be trained differently to be successful, fitting comfortably within a longer running narrative that American schools were or had become largely ineffectual. The consequences of the perceived deficiencies of America's schools had taken on new and higher stakes during the Cold War. Schooling was no longer simply a matter of producing good citizens or providing a means of personal advancement—instead, developing the nation's human capital was now a matter of existential economic and military importance (e.g., Tröhler, 2014). This argument provided additional non-moral justification for improving educational equality: failing to develop the talents of all American youth was a waste of one of the country's most valuable resources. During the 1970s, these views about the role of schools, coupled with America's ongoing economic woes helped drive the rapid rise and proliferation of the Minimum Competency Testing (MCT) movement (Resnick, 1980). The MCT movement, which sought to improve school performance and student achievement by requiring students to pass tests to graduate from a particular grade or school, was arguably the country's first nationwide effort at test-based accountability. Thirty-five states adopted some form of MCT by 1980.

Concerns about school performance, economic competitiveness, and global competition were bolstered by the publication of *A Nation at Risk: The Imperative for Education Reform* in 1983, which added new rhetorical heft and policy aims to the national conversation (National Commission on Excellence in Education, 1983). This report asserted that “the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people,” adding that “if an unfriendly foreign power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war” (National Commission on Excellence in Education, 1983, p. 9). Concern about the mediocrity of school performance led *A Nation at Risk* to reject a focus on minimum competency in favor of calling for educational excellence. This excellence would be achieved by increased attention to higher standards, improved curricula, and greater accountability.

The ideas espoused in *A Nation at Risk* provided the blueprint for future educational reform. The call for better, more rigorous curriculum and standards was adopted by several professional organizations. For instance, the National Council of Teachers of Mathematics published *Curriculum and Evaluation Standards for School Mathematics* in 1989, which called for a novel and more conceptual approach to mathematics instruc-

tion (National Council of Teachers of Mathematics, 1989). Similar calls to shift subject emphasis and approach occurred in history, science, and reading. Congress also created the National Council on Education Standards and Testing in 1991, which issued a report titled *Raising Standards for American Education* and endorsed “the adoption of high national standards and the development of a system of assessments to measure progress toward those standards” across the school curriculum (National Council on Education Standards and Testing, 1992, p. 8). These calls for curricular reform were paralleled by legislative efforts aimed at increasing political pressure for improved school performance. Following a summit of state governors in Charlottesville in 1989, the federal government passed a series of bills—Goals 2000: Educate America Act (1994), Improving America’s Schools Act (1994), and NCLB (2001)—that were intended to incentivize states to raise expectations for student achievement, develop new and more rigorous academic standards, and establish a more regular and robust system of assessment (McGuinn, 2006).

Set among these developments, it is easy to imagine that the *Knowing What Students Know* report committee believed there was value and promise in articulating a more sophisticated, research-driven view of assessment. Indeed, the framing of the report as an effort to cast aside outdated approaches to assessment in favor of more ambitious and rigorous ones was perfectly aligned with two decades of rhetoric about educational reform. As *Knowing What Students Know* clearly identifies, a system of reform predicated on using students’ demonstrated knowledge on standardized assessments to guide system-level changes is only as good as the assessments are. *Knowing What Students Know* also incorporated recent developments in research on cognition and learning (e.g., National Research Council, 1999; Resnick & Resnick, 1992), including highlighting the situated nature of understanding and the importance of cognitive schemas in shaping a person’s ability to learn, recall, and apply information in new contexts—both of which underscore the need to rethink the what, how, and when of assessments (Resnick & Resnick, 1992). *Knowing What Students Know* also reflected the view that traditional assessments, such as those that used multiple choice questions and those that required students to recall basic facts without probing their cognitive processes, could never provide adequate information about student learning or competence in a curricular domain. Likewise, the prevailing view was that the inability of existing assessment systems to measure complex knowledge and skills virtually assured that test results could neither speak to the more ambitious elements of newly adopted standards nor provide sufficiently detailed accounts of student learning to guide teaching and instruction. With these prescient concerns in mind, *Knowing What Students Know* sought to chart a different course for America’s system of educational assessment.

## WHAT ARE BALANCED ASSESSMENT SYSTEMS?

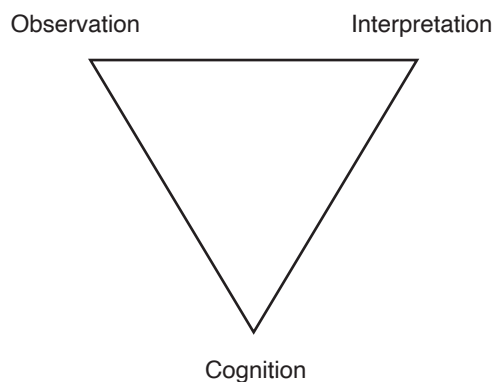
This section focuses on the criteria for describing and evaluating balance in assessment systems as laid out in *Knowing What Students Know*. To build up to these criteria, *Knowing What Students Know* puts forth an argument about the state of assessments in U.S. education at the time of its publication, developing the conceptual justification for balanced assessment systems. Here, we summarize this argument and then present and describe the criteria and how they were operationalized.



*Knowing What Students Know* begins by addressing the nature of assessment and identifies assessment's three main purposes. First, there is *formative assessment*, or assessment to assist teaching and learning. Second, there is *summative assessment*, or assessment to ascertain students' level of competency. Summative assessments can be classroom-based or large-scale, although *Knowing What Students Know* focuses on large-scale assessments given the contemporaneous policy context discussed above. Third, there is assessment to evaluate programs: these are typically based on summative assessments, but instead of using the assessment to make a judgment about an individual, it is used to make a judgment about an institution or policy. The report notes that a single assessment will not be able to serve all these purposes and that there is indeed often misalignment among the various purposes (e.g., that the kinds of assessments useful for teachers' instructional decisions are typically poorly suited for evaluation).

*Knowing What Students Know* argues that assessment is a process of reasoning from evidence. Assessment data (i.e., students' responses to assessment prompts) provide evidence through interpretation. A chain of reasoning helps the author of the assessment determine what to measure and establish a justification for how that measurement produces evidence to address the desired goals of the assessment. *Knowing What Students Know* focuses on the "assessment triangle," which emphasizes three essential elements underlying any assessment—a model of student cognition, a set of beliefs about the kinds of observations that will provide evidence of student competencies, and an interpretation process for making sense of the evidence (see Figure 2-1). The implication of the assessment triangle is that all three components—cognition, observation, and interpretation—must support each other for the assessment to be effective. *Knowing What Students Know* further emphasized that the model of student cognition underlying the assessment triangle should extend to curriculum and instruction.

*Knowing What Students Know* then discusses the state of knowledge on thinking and learning and draws implications for assessment systems, echoing other contemporary accounts of assessment for policy and practice (e.g., Shepard, 2000; Stiggins, 2001). For instance, the report concludes that assessment practices are too focused on component skills and discrete knowledge and not enough on complex aspects of student achievement. The report emphasizes the mind's cognitive architecture and concludes



**FIGURE 2-1** The assessment triangle.  
SOURCE: National Research Council (2001).

that a primary focus of assessment should be on understanding the contents of long-term memory and how people use long-term memory to answer questions and solve problems. It also concludes that assessments should measure important metacognitive skills and problem-solving strategies. By making students' thinking visible, these more advanced forms of assessment can be more instructionally useful to teachers. These forms of assessment also must consider important contextual factors such as students' background knowledge, the context in which the tasks are presented, and the degree of transfer required for success in the task. Importantly though, the report also concludes that the body of evidence on creating these more advanced assessments is insufficient for practical implementation in most cases, requiring further translation to be useful in practice.

*Knowing What Students Know* then proceeds to describe a vision for a modern assessment system, beginning with the importance of a model of cognition and learning for assessment development. For instance, the report argues that if one is assessing to inform arithmetic instruction, test developers need to start from an understanding of how students learn in the tested domains. This model of cognition and learning should be based on empirical research in the domain, be able to differentiate between the performances of novice and expert learners, and account for variation in student learning pathways. The model should also inform assessment construction by identifying appropriate aspects of the larger theory of cognition and learning, and should lend itself to being aggregated for a variety of assessment purposes. *Knowing What Students Know* notes that there will be content areas where models of student learning are not well developed but argues that the general principles articulated above should still hold. The report goes on to discuss examples of models and their application, as well as principles of task and assessment construction for this new model and provides validation and reporting recommendations.

### **The Principles of Balanced Assessment Systems**

With this view of assessment articulated, *Knowing What Students Know* introduces and describes the principles of balanced assessment systems. After describing features of classroom-level and large-scale assessments independently, the report advocates balancing these two forms of assessment, claiming that the status quo at the time of publication was heavily tilted toward large-scale uses. The report then introduces three principles that characterize balanced assessment systems: *comprehensiveness*, *coherence*, and *continuity*. As we discuss below, the decision to introduce principles to define balanced assessments—rather than elements or features—means that determining whether a system has achieved balance is less a categorical determination than one of degree. The challenge posed by fulfilling these principles is both technical and political. For instance, how, to what degree, and with what balance these principles should be pursued would certainly spur debate among experts. Whether the ensuing compromise would be acceptable to the public or feasible in practice given the limited time, resources, and technical expertise available in individual states, school districts, or schools is a matter likely to produce more compromise and perhaps deviation from the ideal.



With this in mind, we present the principles of balanced assessment systems and their definitions, discussing each principle in some detail with an eye toward helping readers think of them in terms of a continuum.

### *Comprehensiveness*

The first principle of balanced assessment systems is comprehensiveness. *Knowing What Students Know* characterizes comprehensiveness by explaining:

*A range of measurement approaches should be used to provide a variety of evidence to support educational decision making. Educational decisions often require more information than a single measure can provide. As emphasized in the NRC report High Stakes: Testing for Tracking, Promotion, and Graduation, multiple measures take on particular importance when important, life-altering decisions (such as high school graduation) are being made about individuals. No single test score can be considered a definitive measure of a student's competence. Multiple measures enhance the validity and fairness of the inferences drawn by giving students various ways and opportunities to demonstrate their competence. The measures could also address the quality of instruction, providing evidence that improvements in tested achievement represent real gains in learning (NRC, 1999c).... Further, in a comprehensive assessment system, the information derived should be technically sound and timely for given decisions. One must be able to trust the accuracy of the information and be assured that the inferences drawn from the results can be substantiated by evidence of various types. The technical quality of assessment is a concern primarily for external, large-scale testing; but if classroom assessment information is to feed into the larger assessment system, the reliability, validity, and fairness of these assessments must be addressed as well. (National Research Council, 2001, pp. 253–255, italic added by authors for emphasis)*

This principle emphasizes the benefits of employing multiple assessment measures, especially in high-stakes instances. This emphasis is in direct response to prevailing uses of assessment in the years prior to *Knowing What Students Know*, especially in high stakes situations (e.g., minimum competency tests, exit exams used to award high school diplomas). *Knowing What Students Know* offers an example of a comprehensive assessment used in the United Kingdom for A-level physics, which combines multiple short sit-down assessments that include a variety of item types with laboratory exercises and essays.

According to *Knowing What Students Know*, comprehensiveness has several benefits. First, more comprehensive assessment systems provide more information than a single measure. They also enhance the validity and fairness of the inferences drawn from the data, so are therefore more trustworthy for users. Finally, more comprehensive systems may also be more instructionally valid (i.e., more useful for discerning effective and ineffective instruction). But to achieve these benefits, the comprehensive assessments must be technically sound and delivered in a timely manner, for both classroom and large-scale assessments. *Knowing What Students Know* also briefly acknowledges that comprehensiveness requires greater cost and effort in terms of assessment development, validation, and scoring.

## Coherence

The second principle of balanced assessment systems is coherence. *Knowing What Students Know* defines coherence as follows:

One dimension of coherence is that the *conceptual base or models of student learning* underlying the various external and classroom assessments within a system *should be compatible*. While a large-scale assessment might be based on a model of learning that is coarser than that underlying the assessments used in classrooms, the conceptual base for the large-scale assessment should be a broader version of one that makes sense at the finer-grained level (Mislevy, 1996). In this way, the *external assessment results will be consistent with the more detailed understanding of learning underlying classroom instruction and assessment*. As one moves up and down the levels of the system, from the classroom through the school, district, and state, assessments along this vertical dimension should align. As long as the underlying models of learning are consistent, the *assessments will complement each other rather than present conflicting goals for learning*.

To keep learning at the center of the educational enterprise, assessment information must be *strongly linked to curriculum and instruction*. Thus another aspect of coherence, emphasized earlier, is that *alignment is needed among curriculum, instruction, and assessment* so that all three parts of the education system are working toward a common set of learning goals. Ideally, assessment will not simply be aligned with instruction, but *integrated seamlessly into instruction* so that teachers and students are receiving frequent but unobtrusive feedback about their progress. If assessment, curriculum, and instruction are aligned with common models of learning, it follows that they will be aligned with each other. This can be thought of as alignment along the horizontal dimension of the system.

To achieve both the vertical and horizontal dimensions of coherence or alignment, *models of learning are needed that are shared by educators at different levels of the system, from teachers to policy makers*. This need might be met through a process that involves gathering together the necessary expertise, not unlike the approach used to develop state and national curriculum standards that define the content to be learned. But *current definitions of content must be significantly enhanced based on research from the cognitive sciences*. Needed are user-friendly descriptions of how students learn the content, identifying important targets for instruction and assessment (see, e.g., American Association for the Advancement of Science, 2001). (National Research Council, 2001, pp. 255–256, italic added by authors for emphasis)

Later reports explicitly distinguish and define vertical and horizontal coherence using the same concepts articulated in the excerpt above (e.g., National Research Council, 2006; Shepard et al., 2018). In the vertical dimension, a common model of student learning helps ensure that different forms and levels of assessment provide complementary, rather than conflicting, information. This definition serves as a rejoinder to large-scale assessments that were seen as poorly aligned with classroom assessments, sending teachers unclear signals about student performance and their own instructional needs. In the horizontal dimension, alignment among—or integration of—curriculum, instruction, and assessment helps ensure relevance and utility of assessment results.

The last paragraph in the definition excerpted above emphasizes the importance of learning progressions—models of student learning of content that are based in cognitive science and are widely shared and understood across the levels of the system. *Knowing What Students Know* notes that existing definitions of content, including those found in standards documents, are insufficiently linked to detailed conceptions of how students learn. In its conclusion, the report advocates for a substantially expanded research agenda to develop and test new conceptual models of student learning.

### *Continuity*

The third and least-developed principle is continuity, which is defined in *Knowing What Students Know* as follows:

In addition to comprehensiveness and coherence, an ideal assessment system would be designed to be *continuous*. That is, assessments should *measure student progress over time*, akin more to a videotape record than to the snapshots provided by the current system of on-demand tests. To provide such pictures of progress, *multiple sets of observations over time must be linked conceptually so that change can be observed and interpreted*. *Models of student progression in learning* should underlie the assessment system, and tests should be designed to provide information that maps back to the progression. (National Research Council, 2001, pp. 256–257, italic added by authors for emphasis)

This principle, reflecting developments in the science of learning and cognition, emphasizes models of student learning, and especially the longitudinal and temporal nature of that learning. Continuity argues for the centrality of assessments for measuring growth, as opposed to the typical one-time assessment practice. It is unstated, perhaps because this report predates the modern “value-added” movement that emphasizes the attribution of growth in student achievement to individual schools and teachers, but the implication is that the growth focus of the continuity principle refers to more than simply a statistical analysis of performance over time. Rather, the focus is on developing longitudinal models of student learning, as well as focusing assessment and reporting on student performance over time against those longitudinal models.

### *How Much Balance Is Enough?*

One of the central tensions we return to throughout this chapter, and that we believe contributes to the difficulties in widely implementing balanced assessment systems, is the lack of clarity about how to measure balance and determine when there is sufficient balance in the system. How much balance is “enough?” As the principles above illustrate, there is no bright line or checklist that says if your system contains X, Y, and Z elements, it is sufficiently balanced.

The three principles of balanced assessment systems and the other criteria discussed later in this chapter are continuous, not dichotomous. For an assessment director seeking to create balance, the task can seem Herculean because more and different forms of assessment can always be added to make the system more comprehensive. One can always tighten the link between assessment and curriculum to bring about more coherence. And one can always add additional longitudinal measurements to deepen

continuity. In contrast to the ever-expanding possibilities of bringing “more balance” are the very real time and resource constraints for developing and implementing assessment systems. Yet, despite this challenge, we are unaware of any clear processes or guidelines for how to create sufficient comprehensiveness, coherence, and continuity. We return to this issue later, and we also note that this challenge of how much is enough is not unique to the effort to achieve balanced assessment systems. Modern argument-based conceptions of validity (Kane, 1992) also suffer from a similar challenge: how much evidence is enough to make a particular validity determination is in the eye of the beholder. Accepting that reasonable minds will differ on the technical matter of how to pursue and balance these principles in the design of an assessment system, we expect that implementation of these systems in practice would foster still more variation, further underscoring the need to assess balance along a continuum.

### **Updates to the Balanced Assessment Criteria**

Over time, the balanced assessment criteria have been expanded and revised in various ways (Marion et al., 2019b; National Research Council, 2003, 2006, 2014). The 2003 NRC report *Assessment in Support of Instruction and Learning: Bridging the Gap Between Large-Scale and Classroom Assessment* expanded significantly on *Knowing What Students Know*, discussing the three criteria alongside two additions to the list: (1) integrated and (2) high-quality assessments (National Research Council, 2003). The 2003 report offered summaries of innovative systems at the time across the United States and the world, but also noted the lack of evidence for the effectiveness of these programs—a problem that persists today. The report also highlighted that “with a few exceptions, little effort has yet been made to transfer these programs to other settings with different characteristics” (National Research Council, 2003, p. 42).

In 2006, the NRC followed up with a deep dive into issues related to state assessment systems for science in *Systems for State Science Assessment* (National Research Council, 2006). This report also emphasized aspects of balance, focusing on horizontal, vertical, and developmental coherence in assessment systems—embodying some elements of coherence, continuity, and comprehensiveness, although using somewhat different terminology. The 2006 report helpfully described different models by which states could meet the NCLB science assessment requirements, while also raising thorny technical considerations.

By 2014, with Next Generation Science Standards (NGSS) becoming more widely adopted and the need for a more coherent approach to science assessment becoming more clear, the NRC issued a report on science assessment systems, *Developing Assessments for the Next Generation Science Standards* (National Research Council, 2014). This report brought together many of the ideas and examples discussed in previous NRC reports and included specific examples of curriculum-embedded tasks that embody the 2014 report’s vision of NGSS-aligned assessment systems. The 2014 report argues that proper assessment of the NGSS *requires* elements of balance, for instance claiming that tasks must “be designed so that they can accurately locate students along a sequence of progressively more complex understandings of a core idea and successively more sophisticated applications of practices and crosscutting concepts” (i.e., continuity, National Research Council, 2014, p. 45). The 2014 report also emphasizes the need

for a systems approach to assessment of the NGSS, including classroom assessments, monitoring assessments, and indicators of opportunity to learn, while noting the difficulties associated with the lack of common curricula across jurisdictions.

Given these developments in the years since *Knowing What Students Know*, recent efforts to define balanced assessment systems introduced two additional criteria: utility and efficiency. Chapter 1 of this volume, “Reimagining Balanced Assessment Systems: An Introduction,” adds another related criterion: a focus on ambitious and equitable teaching and learning, but we do not repeat that argument and explanation here.

### *Utility*

Marion and colleagues (2019b) provide definitions of both utility and efficiency, citing their work with states to operationalize balanced assessment systems. Utility is defined as follows:

*Utility is the degree to which the assessment system provides the information necessary to support its multiple and often diverse purposes. Utility is not evaluated in the abstract, but follows from a well-articulated theory of action specifying the system’s intended outcomes and the processes and mechanisms by which these outcomes are realized (e.g., Hall, 2015). To be sure, assessments are validated for specific purposes and uses. But when considering utility, we must reach beyond the score inferences that are the focus of validity evaluations and rely on a theory of action that spans all of the components of the system. With assessments purportedly designed to improve learning and teaching, these aims often include: providing feedback for identifying and adjusting misunderstandings, promoting deeper learning, fostering student engagement, and/or enhancing self-regulation or/and related skills. Thus, utility should be evaluated by examining the extent to which each assessment experience, and the system as a whole, supports the overarching aims. (Marion et al., 2019b, p. 5, italic added by authors for emphasis)*

This conception of utility represents another bold addition to the already ambitious balanced assessment framework. It reinforces the idea that balanced assessment systems require a coherent theory of student learning and organizational change. Importantly, this definition of utility operates at both the level of individual assessment experience and the whole system of assessment. Given the decentralized structures governing assessment in American education, we argue that utility should be evaluated at the individual district level.

While utility is to be evaluated against each individual system’s theory of action, the definition sets a high bar by implying that assessment and assessment systems should not only improve teaching and learning but also “[provide] feedback for identifying and adjusting misunderstandings, [promote] deeper learning, [foster] student engagement, and/or [enhance] self-regulation or/and related skills” (Marion et al., 2019b, p. 5). If assessment systems are struggling to only improve teaching and learning, these same systems will surely struggle to achieve more complex goals. We note that while the term “utility” was not emphasized in Chapter 1 of this volume, “Reimagining Balanced Assessment Systems: An Introduction,” the goal of fostering ambitious and culturally relevant instruction sets a target for the utility of district assessment systems.



## *Efficiency*

Recognizing both the potentially boundless scope of balanced assessment systems and the growing anti-test political context of the latter half of the 2010s, a final criterion for balanced assessment was added—efficiency—which is defined as follows:

By this we mean *getting the most out of assessment resources and eliminating redundant, unused, and untimely assessments*. Efficiency determinations identify and reduce assessments that are not serving the stated purposes or are redundant with other, more useful assessments. (Marion et al., 2019b, p. 7, italic added by authors for emphasis)

Efficiency is a valuable criterion to thwart the “yes, and” approach to balancing an assessment system. That is, this criterion specifically forces local actors to consider whether each individual assessment in the system is necessary or superfluous, rather than simply piling new assessments on top of existing ones to achieve balance. Efficiency is also especially useful when combined with utility since efficiency focuses mainly on quantity and utility focuses on the quality of each assessment and its alignment with a theory of change.

## **THE IMPLEMENTATION OF BALANCED ASSESSMENT SYSTEMS OVER TIME**

The assessment principles articulated in *Knowing What Students Know* reflect a new era of thinking about the role, design, and implementation of standardized assessments. Having better, more sophisticated ideas is, however, not enough to ensure their faithful adoption or implementation, even when they are backed by research and released under the auspices of an esteemed group like the NRC. Two decades later, there has been only modest success at sustained implementation of balanced assessment systems at scale. Even so, there have been several attempts to pilot new assessment systems and reorient existing systems in ways that reflect the assessment principles in *Knowing What Students Know* (in line with the observation above that balance is a matter of degree, not a bright line). In this section, we review some of these implementation efforts and consider the lessons that can be gleaned by these examples.

### **New Hampshire’s Performance Assessment of Competency Education**

Arguably the most notable example of an effort to reform an assessment system in line with the principles of *Knowing What Students Know* involves a pilot project in New Hampshire. In 2015, New Hampshire took advantage of flexibility provided by the U.S. Department of Education and received a waiver for certain elements of the testing requirements under NCLB to pilot an assessment program called New Hampshire Performance Assessment of Competency Education (PACE) (New Hampshire Department of Education, 2023). The goal of PACE was to develop an assessment system that could better serve the multiple purposes and audiences that utilize such assessment information. While annual state testing regimes developed under NCLB were providing useful information to lawmakers and the broader public, the scores produced by statewide testing were of limited value to classroom teachers in guiding

their instruction. PACE sought to address this shortcoming by developing a multi-layered assessment system involving locally developed and administered performance assessments, common assessments administered across participating districts, and the standard (Smarter Balanced) state-level assessments. Both the local and common tasks used in the PACE system are teacher-designed, and, as a result, are intended to closely resemble classroom tasks and instruction. Common tasks are developed jointly by teachers across districts and are subject to a one-year pilot testing period, during which the performance and scoring of the tasks are assessed for quality and potential biases. Following any necessary revisions, common tasks are administered across participating districts, with the same common task administered across districts in the specified grade. Local tasks are developed by individual teachers or schools, intended for local use, and are not required to undergo the same piloting or evaluation process as the common tasks. Teachers developing local tasks are encouraged to first participate in the development of a collaborative common task so that they have an opportunity to develop their knowledge about and skills for task creation. To reduce the burden on teachers to develop local tasks and to increase the number of high-quality tasks available for use, common tasks that have been piloted and made operational are added to a pool of tasks that can be drawn on by teachers and schools for use as local tasks. Depending on the grade and subject matter, local and common assessments developed under PACE are used to make the required annual state determination of student competency (Becker et al., 2017; Evans & Lyons, 2017).

The PACE pilot closely adheres to the two *Knowing What Students Know* principles of coherence and comprehensiveness. By incorporating teacher-developed competency tasks in addition to traditional standardized assessments, PACE measures of competency involve a much larger variety of tasks and, in turn, can assess a larger array of student skills and abilities than a traditional assessment system. Likewise, because the required performance assessments are so closely linked to classroom curriculum and instruction, the information provided by these assessments is much more likely to provide teachers with information that can inform their planning and instruction. This information will likely prove more useful to teachers because they are both more familiar with the local and commonly administered tasks than they would be with a standardized assessment devised by a third party but also because, at least theoretically, the preparation for and administration of the assessments provides teachers with real time feedback on their instructional practice. As the authors of a formative assessment of PACE explained, “PACE is ... intended to influence instructional practices” but unlike with traditional assessments, “PACE leadership is not overly concerned about teachers ‘teaching to the test.’ PACE, ideally, supports ‘testing to what is taught’” (Becker et al., 2017, p. xxii).

Although PACE is useful in demonstrating how the principles of *Knowing What Students Know* can be used to devise an annual statewide assessment system, the pilot project also illustrates the considerable challenges and resources necessary to implement such a system. Across the first four years of the pilot, only 14 of the state’s 84 school administrative units had implemented the PACE system (Lyons et al., 2017). To participate fully in the pilot (i.e., use PACE across all available grades and subjects), districts had to commit teacher time to developing, piloting, administering, and scoring the local and common assessments required by the PACE system. Given that

full participation implicitly assumes teachers—across grade levels and subject matter expertise—have the ability to develop high quality performance tasks and assess them reliably, participation requires considerable investment in professional development for these teachers.

New Hampshire developed a three-tier system that allowed districts to build this capacity over time with appropriate state level support. Tier 3 districts, or those that had few or no classrooms implementing competency-based learning and personnel with little or no experience developing task-based performance assessments to evaluate competencies, were provided access to school-level consultants to develop local competency targets based on the state’s model standards. Tier 2 districts, or those that had developed the necessary school and course level competency targets and had some—or uneven—experience developing task-based performance assessments, received access to professional development from experts on topics ranging from creating performance tasks to developing reliable scoring procedures to fostering professional learning communities. Tier 1 districts, or those that were fully implementing PACE, were provided expert consulting and coaching assistance as teachers engaged in the multi-step process of developing, piloting, revising, and implementing high quality performance tasks. Tier 1 districts also, with state assistance, invested in more advanced training for select teacher leaders focused on advanced performance assessment, including validity theory and principled assessment design; depth of knowledge, to assist with developing more cognitively demanding assessment tasks; and professional community development, to facilitate collective development of assessment tasks. Finally, teachers from Tier 1 districts were required to attend sessions during the summer where student work on PACE common tasks was discussed and scored (Lyons et al., 2017).

Producing comparable scores is an especially important component of the PACE pilot, given the goal of using local and common tasks in place of statewide assessments for some grades and subjects (Evans & Lyons, 2017). To be viable for these purposes, student scores on PACE tasks across districts needed to support the same inferences about students’ knowledge and skills in a domain. Specifically, the state needed to develop scoring procedures—and processes to monitor the scoring—that would ensure that a student reported proficient on a task in one district would be rated proficient in another district. To ensure this contiguity, samples of student work must be scored by multiple teachers within and across districts to ensure inter-rater reliability and comparability of scores within and between districts. As the multiple evaluations of the PACE program cited throughout this section have found, completing these scoring processes and achieving acceptable levels of scoring reliability—the target is 60 percent—is time intensive. As one study concluded, “the practicality and feasibility of scaling up the proposed methods in a large-scale performance assessment program is a real concern particularly within a state that has many more districts or other units with a large number of different local assessment systems” (Evans & Lyons, 2017, p. 31).

A formative evaluation of PACE, likewise, found that the amount of time required of teachers to develop assessments and calibrate and score student work was an “ongoing source of remaining tension” in the pilot implementation (Becker et al., 2017, p. 49). One-quarter of the teachers surveyed disagreed or strongly disagreed with the statement that the time required by PACE was worth the benefits (Becker et al., 2017). One attempted solution was to schedule a task planning session for the weekends. While



this shift did not reduce the amount of time required, it did reduce the amount of time teachers had to be absent from their classrooms. Some districts have also experimented with reducing the amount of time required to score student work by eliminating scoring calibration sessions and shifting the work of scoring from the school year to the summer. The result, however, was that these districts had substantially lower inter-rater reliability, with the districts failing to achieve the 60 percent target in multiple grades and subjects (Becker et al., 2017). As a result, the pilot protocols were modified to require that calibration sessions and scoring sessions occur within school districts during the school year (National Center for the Improvement of Educational Assessment, 2020).

### Learning Progressions

Another particularly prominent effort to implement elements of balanced assessment systems since *Knowing What Students Know* has been in the area of “learning progressions,” sometimes also called “learning trajectories” (see Corcoran et al., 2009, for an early history; see Shepard, 2018, for a more recent one). Early advocates for learning progressions described their potential benefits as an alternative to traditional ways of thinking about standards, curriculum, assessment, and instruction (Corcoran et al., 2009). Through making and rigorously testing hypotheses about how children develop mastery in core concepts, learning progressions pair nicely with—and require—more balanced forms of assessment (National Research Council, 2007). That is, to assess student progress along a learning progression, one must employ an assessment system that is coherent, comprehensive, and continuous. Indeed, early conceptions of learning progressions emphasized their potential for promoting “[c]learner ties to instruction,” “[providing] reference points for assessments that report in terms of levels of progress,” and “[informing] the design of curricula that are efficiently aligned with what students need” (National Research Council, 2007, p. 9)—all clear nods to the principles of balanced assessment.

The concept of learning progressions has been influential, achieving a foothold in mathematics education (e.g., Clements & Sarama, 2020; Daro et al., 2011) and informing the design of the NRC’s *A Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (2012), which itself informed both the NGSS and other sets of state standards that have been adopted nearly nationwide (Duncan & Rivet, 2013). Shepard (2018) offers a thorough recent review of learning progressions, their development, and their impact. Shepard and colleagues (Shepard, 2018; Shepard et al., 2018) argue that learning progressions are best built “from the bottom up, focusing on local jurisdictions or curricular projects, where it is more likely to be possible to design for coherence among curriculum, instruction, assessment, and teacher learning” (Shepard, 2018, p. 167). Learning progressions have been built, primarily funded by the National Science Foundation, in domains such as matter and atomic-molecular theory, scientific argumentation, modern genetics, energy, evolution, and celestial motion. Shepard (2018) also notes the “extensive research and development and detailed materials” (p. 168) needed to deploy learning progressions, pointing to successful examples like the *Building Blocks* mathematics curriculum that embodies these principles.

Despite some promising small-scale findings, evidence of teaching and learning growth from learning progressions is modest. While there has been substantial funding

allocated toward learning progressions projects through the National Science Foundation, the impact at scale is limited. Most of the research described by Shepard and colleagues (2018) is small in scale, and besides influencing the development of recent standards, there is very little evidence of impact on teaching and learning beyond these controlled studies.

### **Smarter Balanced Assessment Systems**

The Smarter Balanced Assessment was one of the two major Common Core–aligned testing consortia funded through the Obama administration’s Race to the Top program in 2009 (The White House, n.d.). Although Common Core and its assessment consortia have become entangled in political battles across the United States since their initial rollout, as of 2023, 12 states are still Smarter Balanced members (Smarter Balanced, n.d.). While Smarter Balanced is best known for its state summative assessment, the organization also offers elements of its overall design to facilitate better balance in state and local assessment systems. For instance, Smarter Balanced offers partner states optional interim assessments in the form of interim assessment blocks, or brief assessments focused on just a few assessment targets, as well as interim comprehensive assessments that are built to align with and provide scores on the standard Smarter Balanced scale (Hardoin et al., 2020).

These interim assessment systems contain elements that would seem to enhance the balance of local assessment systems as compared to more traditional commercial interim assessments, which often have limited or uncertain alignment with larger summative tests. For instance, the Smarter Balanced interim assessments include a variety of item types. While the constructed response items offered must be scored locally, Smarter Balanced offers training to teachers on scoring. The interim assessments provide results to teachers rapidly and in a form that can be instructionally useful for informing remediation and differentiation. Smarter Balanced also offers Tools for Teachers, aimed at aligning to interim assessments and differentiating instruction based on student performance. According to an independent evaluation of Smarter Balanced in California, evaluators found some evidence that these reforms are widely used and well received in local districts. Where educators found the Smarter Balanced interim assessments less useful was generally where the assessments were mandated in a fixed schedule that did not align well with local curriculum coverage and pacing (Hardoin et al., 2020).

### **Other Examples in Practice**

There are many other examples of innovations in assessment over the 20-plus years since *Knowing What Students Know* that could be said to embody elements of balanced assessment systems. Conley (2018) provides an overview of some of these examples, including case studies of a few districts. Conley does not use the term balanced in his book, but the systems he describes often include principles of balance. For a fuller treatment, we encourage interested readers to read Chapter 6 of *The Promise and Practice of Next Generation Assessment*, but note that the description of implementation and impact of the examples he cites is relatively thin (Conley, 2018).

One type of system Conley (2018) cites is a commercial curriculum system leading to a specialized diploma. He cites several similar examples of such systems, including International Baccalaureate, Cambridge International, and the AP Capstone program. These programs emphasize tight coherence among curriculum and assessment, as well as a developmental approach to teaching and measuring student learning. These systems also include multiple forms of assessment that are both interim and summative.

Another type of balanced assessment system is represented by district–researcher collaboratives under the umbrella of the Assessment for Learning Project (ALP). Sponsored by the Hewlett Foundation, the ALP was focused on “deeper learning” and often funded district collaboratives (Conley, 2018). Again, the projects funded under the ALP emphasized principles related to balanced assessment systems, such as greater emphasis on the instructional utility of assessments, a focus on growth in student mastery, and a comprehensive assessment approach based in multiple measures of student performance. Conley (2018) also provides a case study of the Summit Public Schools, which has worked to better integrate curriculum and assessment and, in collaboration with technology providers, more carefully monitor student growth.

#### **FACTORS THAT HAVE HINDERED THE GROWTH AND IMPACT OF BALANCED ASSESSMENT SYSTEMS**

Although the vision contained in *Knowing What Students Know* was inspiring enough to spur engagement from a generation of scholars, readers would be forgiven for asking whether, this intellectual engagement notwithstanding, balanced assessment systems have ever really “caught on.” Struck by the seeming disconnect between continual intellectual ferment and seemingly modest impact in practice, in this final section, we ask “What went wrong?”—or perhaps more accurately, “What continues to go wrong?” Balanced assessment systems are intuitively appealing, and the ideas underlying these systems seem as if they would be widely supported, both by experts in the assessment and policy fields and by practicing educators. But the accomplishments associated with balanced assessment systems to date are vanishingly modest. Why is it that balanced assessment systems have failed to take hold across states and districts? What needs to change, moving forward, to realize better results?

In this section, we introduce key factors we think have contributed to difficulties in achieving balanced assessment systems in practice, echoing and expanding on an earlier analysis by Marion and colleagues (2019a) and the brief discussion in Chapter 1 of this volume, “Reimagining Balanced Assessment Systems: An Introduction.” Here, we survey both the positive and the negative—what we think is needed to implement a balanced assessment system as compared with what currently exists. This section focuses mainly on large-scale state assessments—the “elephants in the room”—because they, due to federal policy requirements, still drive assessment policy and practice nationwide, and because the extant literature overwhelmingly focuses on state tests. Our recommendations are intended to inform the discussions in the rest of the chapters in this volume, so that the appealing ideas underlying balanced assessment systems might begin to take root in state and district assessment systems. We grouped the challenges into two categories: (1) technical and (2) political and practical (see Table 2-1).

**TABLE 2-1** Factors Challenging Growth of Balanced Assessment Systems

Technical Challenges	Political and Practical Challenges
<ul style="list-style-type: none"><li>• Measuring Multiple Complex Domains</li><li>• Interpreting Information Across Grade Levels for Multiple Dimensions</li><li>• Weighting Multiple Measures</li><li>• Scoring Student Work</li><li>• Implementing and Adapting Technology</li></ul>	<ul style="list-style-type: none"><li>• Poorly Designed Assessment and Curriculum Policies</li><li>• Shifting Political Barriers</li><li>• Challenges of Embedding Assessments in the Curriculum</li><li>• Lack of Capacity Across Levels of the System</li><li>• Instructional Reform in the Context of Loosely Coupled Systems</li></ul>

### Technical Challenges

As we have hinted throughout this chapter, obtaining balance in assessment systems is technically difficult for a variety of reasons. Conley (2018) identifies some of the most important technical challenges in next generation assessment. He does not, like many researchers working in this space, use the term balance, although the principles he describes are similar to those in balanced assessment systems. We have not repeated his treatment of this topic here, but briefly summarize and elaborate on key technical issues that have challenged efforts to develop balanced and next generation assessments alike. Experts seeking to address these technical challenges are likely to have differing views or arrive at alternative conclusions of how to manage them, depending on the specific context. Given the complexity involved, it is unlikely that districts will be able to find off-the-shelf assessments that are sufficiently curriculum-embedded to create a balanced system. More likely, districts would need to retain and consult experts as they design an assessment system with the desired level of balance or attempt to adapt an off-the-shelf option.

#### *Measuring Multiple Complex Domains*

Typically, standardized assessments focus on one content domain and, to a large extent, one or two levels of cognitive complexity—mostly memorization and procedures (see, e.g., Polikoff et al., 2011). Because they aim to capture a more authentic picture of what students know and can do, balanced assessment systems must measure complex thinking skills, and creating these measures is simply more technically challenging than the more traditional item types. Consider the NGSS and their three-dimensional structure—disciplinary core knowledge, crosscutting concepts, and scientific practices. The method of constructing and scoring a set of items measuring all three of these dimensions is no small task.

#### *Interpreting Information Across Grade Levels for Multiple Dimensions*

Balanced assessment systems will typically seek to make more productive use of the longitudinal nature of assessment than traditional systems, including by using vertical scales that span grades. Vertical scales have their own well-established challenges (see, e.g., Briggs, 2013). These challenges are only further compounded by the greater ambition of balanced assessment systems to measure more complex domains. Put simply,

it is hard enough to create a vertical scale for mathematics skills, but what does it look like to create one for the ability to solve complex real world problems?

### *Weighting Multiple Measures*

In balanced assessment systems with multiple measures, scores must be aggregated. This is especially true for assessment systems that are used, in whole or in part, in accountability systems. Conley (2018) discusses the example of California’s CORE districts and their attempts to combine multiple measures through complex weighting. Setting weights in these systems is typically more art than science, and relies heavily on value judgments from the practitioners and policymakers who seek to use the data for various purposes.

### *Scoring Student Work*

Balanced assessment systems must include more complex tasks to measure more complex skills, as well as to make the results more instructionally useful to teachers—although we note that many would argue that state summative assessments can never be useful to teachers because they cannot inform classroom instruction, full stop. But more complex tasks are dramatically more onerous to score than simple tasks, as well as to validly score across classrooms and schools and in ways that teachers can use the resulting information to diagnose and address student learning needs—regardless of whether those complex tasks are part of a local or state level assessment. Conley (2018) discusses some strategies that can help with the scoring burden. One strategy is for teachers to eschew some of the “busy work” they would usually assign and score, which would allow them to focus their energies on the admittedly larger charge of scoring more complex tasks. Another strategy is for teachers to involve students in scoring—for instance, students giving each other peer feedback before the final due date. In this case, teachers can provide high-level feedback to the class based on examining all the students’ work, but only provide scores for individual students. Schools can also structure schedules to facilitate the grading of more complex work. Conley nods to the complexity of this work but argues that

scorers can achieve high levels of agreement ... when they are properly trained in the use of a scoring guide, ... use criterion-based decision-making processes, and are well trained on exemplars so that they ... are able to apply mental models of what they are looking for. (Conley, 2018, p. 157)

While this is undoubtedly the case, finding the time and resources to properly train teachers on scoring guides and exemplars, as well as establishing and maintaining high levels of agreement given the inevitable turnover in personnel, poses another obstacle for districts and schools seeking to implement balanced assessments.

### *Implementing and Adapting Technology*

Finally, Conley (2018) discusses the opportunities and limitations of technology in the context of next-generation assessment. Certainly, technology can offer advantages



in terms of administering and scoring more complex item types. Technology can also aid in producing scores more rapidly and presenting student performance in ways that are more usable to teachers. However, both teachers and students need to be able to use the technology in order for these opportunities to materialize, and that is far from a given. Additionally, some of the best technological innovations will come from smaller startup education technology companies, but Conley (2018) notes that these companies are often disadvantaged in procurement processes.

### **Political and Practical Challenges**

Even if these technical challenges could be overcome at a large scale, there are also important political and practical barriers that impede more balanced assessment systems from taking root in districts and states.

#### *Poorly Designed Assessment and Curriculum Policies*

Balanced assessment systems are complicated to enact, and without incentives and support it is unlikely that implementation will involve anything more than isolated instances of local implementation. Unfortunately, there have been few incentives to develop assessment systems in line with *Knowing What Students Know*. Although it is worth noting that the New Hampshire pilot discussed above was facilitated by the U.S. Department of Education's NCLB waiver process and now the Every Student Succeeds Act (ESSA) authorized, Innovative Assessment Demonstration Authority. Far from encouraging the implementation of balanced assessment systems, there are a variety of ways in which state and federal assessment and curriculum policies currently undermine their spread.

As an example, consider the ways that federal assessment peer review guidance could support balanced assessment systems but instead falls short. Historically, peer review guidance has required that states base accountability decisions on a single summative year-end assessment of student knowledge and skills related to grade-level standards. This requirement makes sense from the standpoint of generating a point-in-time estimate of student proficiency against grade-level standards, but it conflicts directly with the utility of the test results for improving teaching and learning. At best, the results of current accountability assessments could be useful for teachers in the next academic year, but even a passing conversation with practicing educators makes clear that state accountability tests are generally viewed as virtually useless for informing instructional choices. Indeed, many would argue that this is explicitly not the purpose of these assessments, although the messaging around the intended uses of state summative tests is far from clear. This type of requirement also directly contradicts the principles of balanced assessment systems, in that assessments with stakes should be based on a range of types of evidence (i.e., comprehensiveness).

Federal testing and accountability policy under NCLB emphasized the importance of “percent proficient” as the primary metric for school effectiveness. This approach takes the wealth of assessment data available for each test taker, boils it down to a single score, and then dichotomizes that score to either above or below proficient. Under NCLB, growth-based approaches to assessing students or evaluating school



effectiveness were verboten. More recently, under ESSA, federal policy has allowed for growth-based measures of performance but still emphasizes that states must place a heavy emphasis on grade-level proficiency (Every Student Succeeds Act, 2015). These requirements directly contradict the balanced assessment criterion of continuity.

The third pillar of balanced assessments is coherence, and here too federal assessment policy hinders adoption or expansion. Federal assessment policy says nothing of consequence about curriculum, as federal policy intentionally stays far from curriculum issues, and the result is that states vary considerably in their effort—or lack thereof—to ensure students have access to aligned curriculum materials and tightly coupled assessments. Federal policy also does not give so much as a passing nod to theories or models of learning. Theories of learning are not typically emphasized in state standards, although, as is mentioned previously, the NGSS were informed by learning progressions in science.

ESSA is widely seen to offer more opportunities than NCLB for states to improve their assessment systems, moving them more in line with the principles of balanced assessment systems. For instance, ESSA requires states to extend beyond only reading and math proficiency, allows states to use growth-based measures of achievement, and permits more innovative forms of assessment (Conley, 2018). Still, federal requirements and regulations, such as the requirement that every student be tested and receive an individual score indicating their mastery of grade-level content, have substantial impact on states' decisions about the design and implementation of their assessment systems. The result is, despite the modest affordances of ESSA, state assessment systems look like the systems that have been historically required under federal law, not the kinds of systems advocates of balanced assessment systems would prefer.

This argument is not to say that federal policy could not support balanced assessment system principles in practice—quite the contrary. Policy tools exist to encourage or require states to adopt better assessments, but these policy tools are not being used. The most straightforward tool—and the one that the federal government has been most adept at using—is money. But there are other tools as well, including clear and specific guidance, regulation, and enforcement aligned with balanced assessment system principles. For instance, federal policies could encourage innovation in assessment systems, set high bars for the use of assessments for consequential decisions, and encourage states to facilitate tight assessment and curriculum alignment. We return to some of these issues in more detail below, and Chapter 9 of this volume, “Policy Influences on Ambitious Classroom Instruction, Assessment, and Learning,” includes more thoughts on the ways policy can influence instruction and assessment.

### *Shifting Political Barriers*

To create and sustain complex educational reforms like balanced assessment systems, substantial political and structural challenges must be overcome. The political barriers to educational reform have been well described elsewhere (e.g., Polikoff, 2021), but are worth briefly elaborating on here as well. First, and perhaps the defining characteristic of American education, is its decentralization—it includes 50 states and 13,000 school districts, each with their own elected and appointed officials, and each creating and seeking to implement policy (Polikoff, 2021). Without some level of top-

down leadership on assessment issues, wide-scale adoption of balanced assessment system features is likely impossible. However, there is often profound resistance from the public to the perception that states or the federal government are usurping local authority—a seemingly permanent source of tension.

At each level of the decentralized American educational system, political leadership is often unstable, with rapid fluctuations from party to party—and even within a party over time as priorities and goals change. Consider, for instance, the rapid shifts in guidance related to transgender students and Title IX requirements as federal administrations changed during the 2010s (Hersher & Johnson, 2017). Although assessment policy may be somewhat insulated from this instability—it has endured over multiple decades across both Republican and Democratic administrations—it is likely that more ambitious assessment reforms would run the risk of falling victim to political instability of one form or another as happened to the Common Core assessment consortia (Jochim & McGuinn, 2016).

But beyond the mere instability itself, there are the challenges associated with elected or appointed educational leadership positions. The disconnect between rhetorical cycles of educational reform and the time necessary to secure real change is an old problem (Tyack & Cuban, 1997). More often than not, the desire among elected officials is to have short-term political victories, which are usually characterized by a claimed improvement in some type of student outcomes (Tyack & Cuban, 1997). There seems to be little appetite for the sustained, hard work that would be required to build and maintain complex policy instruments like balanced assessment systems. These systems require both infusions of initial capital and sustained resources over time, and are unlikely to produce the short-term bumps in performance that many elected officials would like to be able to point to.

Of course, there are counterexamples that show what is possible with sustained vision and leadership focused on appropriately using available levers of government. For example, Louisiana’s reforms started under State Superintendent John White but have continued for more than 10 years and have substantially revised the state’s approach to curriculum and assessment (Kaufman et al., 2016, 2018). Louisiana built its curriculum reforms around a coherent theory of change, aligning key elements like professional learning around their curriculum-driven vision. The state created powerful incentives for local school districts to adopt and use high-quality curriculum materials, using the power of the Louisiana Department of Education to rapidly encourage adoption. It also provided or identified providers of aligned professional development, leading to sustained teacher learning. When Secretary White stepped down, these reforms had become embedded in Louisiana’s educational culture, and persisted into the subsequent administration. By creating a coherent vision and building a supportive constituency through careful policy design, this approach ensured the longevity of the reforms (Kaufman et al., 2016, 2018).

### *Challenges of Embedding Assessments in the Curriculum*

Curriculum-embedded assessments are at the heart of balanced assessment systems. A balanced assessment system requires tight linkages among assessment, curriculum, and instruction; and, ideally, assessment systems will not merely be aligned

with instruction and curriculum, but seamlessly integrated into instruction. While the goal of curriculum-embedded assessments is admirable, this goal is currently far from the curriculum and instructional reality of American schools.

In most U.S. states, and in many grades and subjects even within the most curriculum-active states, there is very little curriculum centralization (Polikoff, 2021). Many states have no guidance about what curriculum materials schools and districts should adopt. Other states put out lists of approved materials in some subjects and grades, but make those lists strictly advisory. Modest incentives or requirements for districts and schools to adopt particular materials are only offered in a vanishingly small number of state, subject, and grade combinations (Polikoff, 2021). Almost no states keep track of which materials are being used where—and even when they do, the information is often unreliable (Hutt & Polikoff, 2020). What little data we have suggests that there is very little consistency across districts in which materials they adopt (Polikoff, 2021). This fact alone is almost fatal to the idea of state-driven, curriculum-embedded balanced assessment systems—without greater centralization in curriculum decisions, it is hard to see how states can meaningfully support curriculum-embedded balanced assessment systems. To be sure, local actors could still build higher-quality and embedded assessments in their own adopted materials, but this would require substantial capacity—we discuss this possibility below.

Beyond formally adopted curriculum materials, there is the question of how teachers make use of the curriculum materials they are given. Again, the reality of the American educational system is that teachers typically use core curriculum materials as one source among many for instructional guidance. Teachers in U.S. classrooms overwhelmingly engage in various forms of curriculum supplementation (Silver, 2022). Survey data indicate that nearly all teachers supplement with materials from the Internet, from their own repositories, and with materials they create, often with staggering frequency. Curriculum and instruction are indeed the single domain over which individual teachers have the most control (Ingersoll, 2006). Again, the extent of teacher authority over curriculum—and the degree to which teachers exercise that authority by modifying, adding to, or subtracting from the formally adopted curriculum—is something of a stake in the heart of the idea of widespread adoption and implementation of balanced assessment systems.

These realities about curriculum control in U.S. schools and classrooms run headlong into the goal of widespread curriculum-embedded assessments. One path through these challenges is a more assertive state role in curriculum decisions, something that has been advocated and discussed at length elsewhere (Polikoff, 2021). Briefly, this path would include stronger state guidance or requirements for the selection of curriculum from among a small set of high-quality options, coupled with the creation and use of embedded assessments in those same materials. If all districts in a state were using, for instance, one of three highly regarded curriculum materials—and if the state, or the curriculum provider itself, could create and support embedded, ongoing assessment—this could offer a path toward balance. Louisiana’s recent waiver from the U.S. Department of Education to develop a curriculum-embedded assessment system shows what is possible when a state has greater centralization and control over curriculum, although that effort appears too early to have had meaningful evaluation (NWEA, 2021). Louisiana’s approach addresses several of the principles of balanced assessment—most notably, it

is curriculum-embedded (which can only work in a state where substantial proportions of districts use the same materials), drawing on the English language arts (ELA) texts and content that students have used throughout the school year.

But even Louisiana’s approach will run up against the realities of how U.S. teachers use core curriculum materials. The scope and nature of curriculum supplementation is an issue that policy barely attempts to address, but that must be tackled in order for a balanced assessment system to take root. Certainly, there is little interest in meaningfully restricting teachers’ curriculum control, but there may be ways to productively redirect curriculum supplementation in ways that support, rather than undermine, the core curriculum and its embedded assessments. For instance, building collaborative structures and clear expectations that encourage teachers to collaboratively supplement within schools and districts could allow for sufficient between-classroom consistency that would allow balanced assessment systems to become more locally feasible (see Polikoff, 2021, for more discussion of this vision).

#### *Lack of Capacity Across Levels of the System*

Over the past two decades of standards-based assessment, states and districts have developed substantial experience in implementing assessments, using assessment data, and messaging assessment results to families and other stakeholders. But balanced assessment systems are much more complex than the traditional assessment systems they seek to replace. For instance, they require multiple measures, not just one, to make important decisions. They require greater timeliness in reporting—and, simultaneously, more sophisticated forms of evidence from more complex items. They necessitate deeper, shared understanding among educators across classrooms and grade levels, as well as more seamless integration of assessments and their results in the curriculum. In short, they require greater capacity for designing, carrying out, and using assessments across actors in the system.

The implication of implementing balanced assessment systems is that there needs to be substantial assessment capacity in the nation’s educational systems, and if that capacity does not already exist, that effective and ongoing capacity building will take place. However, assessment literacy has always been a sore spot for our educational systems (Popham, 2009). Teacher education programs have historically spent little, if any, time covering assessment literacy (Stiggins, 2006), and there is little evidence to suggest this has changed (Popham, 2018). After *Knowing What Students Know*, researchers like Stiggins (2006) laid out principles for teacher in-service and pre-service education in order to build teachers’ assessment literacy, but these changes to existing protocol have not happened. Without a substantial increase in the assessment capacity of individual educators, achieving the vision of classroom-driven balanced assessment systems advocated in this volume is likely impossible. Chapter 5 of this volume, “Assessment Literacy and Professional Learning,” offers some thoughts on how professional learning can support balanced assessment systems.

There are many reasons for the failure to build assessment literacy across the system, and these are correlated with the issues already discussed in this chapter. Teacher education, both pre-service and in-service, is highly decentralized, with thousands of teacher training programs in operation and very little in the way of standardized expectations.

Teacher educators themselves are often highly resistant to assessment-driven reform (Cochran-Smith, 2006), although they might be more receptive to balanced assessment systems than more traditional forms of test-driven accountability. Finally, in-service teacher learning opportunities are notoriously poor in both design and impact (Darling-Hammond et al., 2017). These are all substantial barriers to overcome.

Perhaps due to the difficulty in achieving assessment literacy through policy, commercial providers have stepped in. For example, large-scale interim assessment providers like the Northwest Evaluation Association and Curriculum Associates provide assessments to thousands of school districts. These assessments can be curriculum-embedded (e.g., in the case of Curriculum Associates, which offers a companion curriculum), but are not necessarily so. The assessments can also be “embedded” in the more vernacular sense of the term, in that they are scheduled to occur at fixed time points during the school year, but not meaningfully embedded in terms of the content they emphasize. Indeed, claims of alignment of interim assessments with curriculum or standards regularly go unverified (Perie et al., 2007). These assessments have met a need that districts had for reasonably high-quality assessments that could be quickly analyzed and used to provide feedback on student progress throughout the year, but they often have fallen far short of contributing to balance in practice.

#### *Instructional Reform in the Context of Loosely Coupled Systems*

The loose coupling (Weick, 1976) that characterizes educational systems in the United States makes complex reform extremely challenging (e.g., Labaree, 2012). Key elements of loosely coupled systems include an absence of regulation, the failure of leaders to influence subordinates, decentralization of power, autonomy of ground-level employees, and a lack of consensus around goals (Weick, 1976). While these characteristics thwart substantial reform efforts, they also can serve advantageous or protective functions, such as allowing the organizations to endure constantly changing environments, permitting failures in some systems without damaging the broader organization, and enabling local adaptation (Labaree, 2012).

Loose coupling has contributed to the standards movement’s lack of success in the last several decades (Polikoff, 2021). Regarding standards-based reforms, states have largely left difficult implementation decisions to local policymakers (e.g., decisions around teacher learning and curriculum adoption). As a result, teachers have almost never received the types of clear guidance needed to understand, let alone implement, complex instructional policies. These challenges have become even more fraught with increasingly complex college- and career-ready standards (Polikoff et al., 2022), which move topics across grades, include more emphasis on conceptual understanding, and often include additional dimensions like mathematics or science practices on top of content expectations. These challenges create an inertia for existing practices that is difficult to overcome.

One way to understand education reform since the 1990s is as various efforts to try and more tightly couple levels of the system, including federal policy to state policy, state policy to student learning outcomes, and state policy to teacher instruction. These efforts have been limited by the factors outlined in this chapter, including shifting politics and policies, the limited capacity in the system, and the increasing ambitions for



our standards and assessment systems (e.g., Cohen et al., 2022). Balanced assessment systems, too, represent a highly ambitious reform at the intersection of assessment and instruction. This chapter’s analysis points to the need to commit to ongoing development of assessment systems while simultaneously working to create tighter couplings in order to see more meaningful implementation.

## CONCLUSION

Who could be opposed to a balanced assessment system? Certainly no one wants imbalance. Yet, most would agree that our assessment systems are currently and have been imbalanced. They were imbalanced when *Knowing What Students Know* was first published, and they are imbalanced today, although perhaps in different ways. The role and quality of state summative tests has ebbed and flowed over time, while the use of interim assessments has exploded since the publication of *Knowing What Students Know*. At the same time, the technology to bring about balance has grown. Advances in assessment quality, spurred in part by the Common Core, have brought better large-scale assessments (Doorey & Polikoff, 2016). Some states have also increased their authority over curriculum materials, making it far more possible for curriculum-embedded assessments to take hold than in a laissez-faire curriculum market.

Still, the nation is far from achieving balance in assessment systems at scale, and the purpose of this chapter is to discuss some reasons for this failure. In terms of lessons learned, we think there are several.

First, achieving balance must be made both more understandable and feasible for educators and local and state policymakers. The criteria underlying balanced assessment systems are laudable, but the ideas are too complex for widespread comprehension and implementation in the current highly decentralized, capacity-poor education systems. Also, even when there is general agreement on the underlying principles, the proliferation of similar ideas with different terminology has added confusion and created the sense that even similar-minded districts are pursuing different paths. It is likely that state departments of education, perhaps working in concert with curriculum developers and providers, must play a larger role in giving local actors clear guidance on how to make assessment systems more balanced. It cannot be “here is the guidance, go forth and conquer;” it must be closer to “here is what you should do, and here are some tools you can use to do it.” Despite the need for this clear guidance, as we have noted throughout, we do not think the literature is clear and specific enough in describing examples of balanced assessment systems and demonstrating their efficacy. One further challenge is that states may lack the necessary capacity—either technical or political—or the will to offer this extra level of support. But acknowledging this problem only underscores the point that this work must be centralized, as these difficulties are only compounded when left to individual districts.

Second, a national policy discussion about the role of state summative assessments in accountability is needed. The status quo presents a situation in which state standardized tests are limping along, supported weakly by many but strongly by few; accountability uses have diminished but educators still feel considerable pressure to tend to their students’ test scores; and state tests are widely pilloried for not providing useful data to inform instruction—a purpose they were never well suited to serve. This



situation serves no constituency well. To do better, an approach to large-scale assessment that ensures appropriate safeguards to protect the rights of underserved students while minimizing the distorting consequences on teaching and learning is needed. Such an approach could also make space for higher-quality and more useful curriculum-embedded assessments that would improve overall balance. Without changes in the state and federal policy context around assessment, balance will be unachievable at lower levels, and the proverbial cart will continue to drive the horse.

Third, assessment experts must be more honest and realistic about the utility of better assessment systems. There is a long history of assessment innovations being oversold, and balanced assessment systems are no different. We believe that well-designed efforts to bring about greater assessment balance would be beneficial, but they would, like all education policy innovations, provide an incremental improvement, not a revolutionary one. They must be coupled with other policies known or strongly suspected to improve student learning, including more generous and equitable school financing; high-quality, highly usable curriculum materials; more and better use of instructional time; and more well-trained educators.

The two decades since the publication of *Knowing What Students Know* have provided ample time for the field to relearn a very old lesson in school reform: describing a better way to do things is never enough to bring about change. Only by tending to the political and organizational demands of reforming ideas can we ever hope to secure a place for them in our schools.

## REFERENCES

- Becker, D. E., Thacker, A. A., Sinclair, A., Dickinson, E. R., Woods, A., & Wiley, C. R. H. (2017). *Formative evaluation of New Hampshire's Performance Assessment of Competency Education (PACE)*. HumRRO. <https://www2.ed.gov/policy/elsec/guid/stateletters/nhpaceformativevalrpt2017.pdf>.
- Briggs, D. C. (2013). Measuring growth with vertical scales. *Journal of Educational Measurement*, 50(2), 204–226.
- Brookhart, S. M. (2013). Comprehensive assessment systems in service of learning: Getting the balance right. In R. W. Lissitz (Ed.), *Informing the practice of teaching using formative and interim assessment: A systems approach* (pp. 165–184). Information Age Publishing.
- Bush, G. W. (2000). *Speech to the NAACP's 91st Annual Convention*. <https://www.washingtonpost.com/wp-srv/onpolitics/elections/bushtext071000.htm>.
- Clements, D. H., & Sarama, J. (2020). *Learning and teaching early math: The learning trajectories approach* (3rd ed.). Taylor and Francis.
- Cochran-Smith, M. (Ed.). (2006). *Policy, practice, and politics in teacher education*. Corwin Press.
- Cohen, J., Hutt, E., Berlin, R., & Wiseman, E. (2022). The change we cannot see: Instructional quality and classroom observation in the era of Common Core. *Educational Policy*, 36(6), 1261–1287. <https://doi.org/10.1177/0895904820951114>.
- Conley, D. T. (2018). *The promise and practice of next generation assessment*. Harvard Education Press.
- Corcoran, T., Mosher, F. A., & Rogat, A. (2009). *Learning progressions in science: An evidence-informed approach to reform*. Center on Continuous Instructional Improvement, Teachers College, Columbia University. <https://files.eric.ed.gov/fulltext/ED506730.pdf>.
- Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute.
- Daro, P., Mosher, F. A., & Corcoran, T. (2011). *Learning trajectories in mathematics: A foundation for standards, curriculum, assessment, and instruction*. Consortium for Policy Research in Education. <https://files.eric.ed.gov/fulltext/ED519792.pdf>.

- Doorey, N., & Polikoff, M. (2016). *Evaluating the content and quality of next generation assessments*. Thomas B. Fordham Institute. <https://files.eric.ed.gov/fulltext/ED565742.pdf>.
- Duncan, R. G., & Rivet, A. E. (2013). Science learning progressions. *Science*, 339(6118), 396–397. <https://doi.org/10.1126/science.1228692>.
- Evans, C. M., & Lyons, S. (2017). Comparability in balanced assessment systems for state accountability. *Educational Measurement: Issues and Practice*, 36(3), 24–34. <https://doi.org/10.1111/emip.12152>.
- Every Student Succeeds Act. 20 USC § 6301. (2015). <https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf>.
- Goals 2000: Educate America Act. 20 USC § 5801. (1994). <https://www.congress.gov/103/statute/STATUTE-108/STATUTE-108-Pg125.pdf>.
- Hardoin, M. M., Dvorak, R. N., Dickinson, E., Paulsen, J., Gribben, M., & Revivo, R. (2020). *California Assessment of Student Performance and Progress (CAASPP): 2020 independent evaluation report (Volume 1)*. HumRRO. <https://www.cde.ca.gov/ta/tg/ca/documents/caaspp20evalrptvol1.pdf>.
- Harris, K. (2020). Forty years of falling manufacturing employment. *Beyond the Numbers: Employment and Unemployment*, 9(16).
- Hersher, R., & Johnson, C. (2017, February 22). Trump administration rescinds Obama rule on transgender students' bathroom use. *NPR The Two-Way*. <https://www.npr.org/sections/thetwo-way/2017/02/22/516664633/trump-administration-rescinds-obama-rule-on-transgender-students-bathroom-use>.
- 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 USC § 2701. (1994). <https://www.govtrack.us/congress/bills/103/hr6/text>.
- Ingersoll, R. M. (2006). *Who controls teachers' work?: Power and accountability in America's schools*. Harvard University Press.
- 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), 44–53.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527–535. <https://doi.org/10.1037/0033-2909.112.3.527>.
- Kaufman, J. H., Cannon, J. S., Culbertson, S., Hannan, M. Q., Hamilton, L. S., & Meyers, S. (2018). *Raising the bar: Louisiana's strategies for improving student outcomes*. RAND Corporation. [https://www.rand.org/pubs/research\\_reports/RR2303.html](https://www.rand.org/pubs/research_reports/RR2303.html).
- Kaufman, J. H., Thompson, L. E., & Opfer, V. D. (2016). *Creating a coherent system to support instruction aligned with state standards: Promising practices of the Louisiana Department of Education*. RAND Corporation. [https://www.rand.org/pubs/research\\_reports/RR1613.html](https://www.rand.org/pubs/research_reports/RR1613.html).
- Labaree, D. F. (2012). *Someone has to fail: The zero-sum game of public schooling*. Harvard University Press.
- Lyons, S., Evans, C., Marion, S., & Thompson, J. (2017). *New Hampshire Performance Assessment of Competency Education technical manual*. National Center for the Improvement of Educational Assessment.
- Marion, S., Thompson, J., Evans, C., Martineau, J., & Dadey, N. (2019a). *The challenges and opportunities of balanced systems of assessment: A policy brief*. National Center for the Improvement of Educational Assessment. <https://files.eric.ed.gov/fulltext/ED598421.pdf>.
- Marion, S., Thompson, J., Evans, C., Martineau, J., & Dadey, N. (2019b). *A tricky balance: The challenges and opportunities of balanced systems of assessments*. Paper presentation at National Council on Measurement in Education, 2019, Toronto, Ontario.
- McGuinn, P. J. (2006). *No Child Left Behind and the transformation of federal education policy, 1965–2005*. University Press of Kansas.
- National Center for the Improvement of Educational Assessment. (2020). *New Hampshire's innovative assessment system: Performance Assessment of Competency Education (PACE): Evaluating technical quality (Volume 2): Results*. <https://www.education.nh.gov/sites/g/files/ehbemt326/files/files/inlinedocuments/pacemanualvol2results.pdf>.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Government Printing Office. [http://edreform.com/wp-content/uploads/2013/02/A\\_Nation\\_At\\_Risk\\_1983.pdf](http://edreform.com/wp-content/uploads/2013/02/A_Nation_At_Risk_1983.pdf).

- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. National Council of Teachers of Mathematics. <https://archive.org/details/curriculumevalua00nati/mode/2up>.
- National Council on Education Standards and Testing. (1992). *Raising standards for American education*. Government Printing Office. <https://files.eric.ed.gov/fulltext/ED338721.pdf>.
- National Research Council. (1999). *How people learn: Bridging research and practice*. National Academy Press. <https://doi.org/10.17226/9457>.
- 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. (2003). *Assessment in support of instruction and learning: Bridging the gap between large-scale and classroom assessment: Workshop report*. The National Academies Press. <https://doi.org/10.17226/10802>.
- National Research Council. (2006). *Systems for state science assessment*. The National Academies Press. <https://doi.org/10.17226/11312>.
- National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K–8*. The National Academies Press. <https://doi.org/10.17226/11625>.
- National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press. <https://doi.org/10.17226/13165>.
- National Research Council. (2014). *Developing assessments for the Next Generation Science Standards*. The National Academies Press. <https://doi.org/10.17226/18409>.
- New Hampshire Department of Education. (2023). *Performance Assessment of Competency Education*. <https://www.education.nh.gov/who-we-are/division-of-learner-support/bureau-of-instructional-support/performance-assessment-for-competency-education>.
- No Child Left Behind Act of 2001. 20 USC § 6301. (2001). <https://www.congress.gov/bill/107th-congress/house-bill/1/text>.
- NWEA. (2021). *Louisiana and NWEA: Creating innovative assessments to foster equity and deeper learning*. NWEA State Solutions. [https://www.nwea.org/resource-center/white-paper/47609/Louisiana-IADA-Brief\\_NWEA\\_whitepaper-1.pdf](https://www.nwea.org/resource-center/white-paper/47609/Louisiana-IADA-Brief_NWEA_whitepaper-1.pdf).
- Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system*. Aspen Institute.
- Polikoff, M. (2021). *Beyond standards: The fragmentation of education governance and the promise of curriculum reform*. Harvard Education Press.
- Polikoff, M. S., Desimone, L. M., Porter, A. C., Garet, M. S., Stornaiuolo, A., Pak, K., Flores, N., Smith, T. M., Song, M., Fuchs, L. S., Fuchs, D., & Nichols, T. P. (2022). *The enduring struggle of standards-based reform: Lessons from a national research center on college and career-ready standards*. EdWorkingPaper 22-622. Annenberg Institute at Brown University.
- Polikoff, M. S., Porter, A. C., & Smithson, J. (2011). How well aligned are state assessments of student achievement with state content standards? *American Educational Research Journal*, 48(4), 965–995. <https://doi.org/10.3102/0002831211410684>.
- Popham, W. J. (2018). *Assessment literacy for teachers in a hurry*. ASCD.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>.
- Resnick, D. P. (1980). Minimum competency testing historically considered. *Review of Research in Education*, 8, 3–29. <https://doi.org/10.2307/1167122>.
- 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 assessments: Alternative views of aptitude, achievement and instruction*. Kluwer Academic Publishers. [https://doi.org/10.1007/978-94-011-2968-8\\_3](https://doi.org/10.1007/978-94-011-2968-8_3).
- Shepard, L. A. (2018). Learning progressions as tools for assessment and learning. *Applied Measurement in Education*, 31(2), 165–174. <https://doi.org/10.1080/08957347.2017.1408628>.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. <https://doi.org/10.3102/0013189X02900700>.
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, 37(1), 21–34. <https://doi.org/10.1111/emip.12189>.

- Silver, D. (2022). A theoretical framework for studying teachers' curriculum supplementation. *Review of Educational Research*, 92(3), 455–489. <https://doi.org/10.3102/00346543211063930>.
- Smarter Balanced. (n.d.). The power of partnerships and collaboration. <https://smarterbalanced.org/our-vision/partnerships>.
- Stiggins, R. J. (2006). Assessment for learning: A key to student motivation and learning. *EDGE*, 2(2), 1–19.
- Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, 20(3), 5–15. <https://doi.org/10.1111/j.1745-3992.2001.tb00065.x>.
- The White House. (n.d.). Race to the Top. <https://obamawhitehouse.archives.gov/issues/education/k-12/race-to-the-top>.
- Tyack, D., & Cuban, L. (1997). *Tinkering toward utopia: A century of public school reform*. Harvard University Press.
- Tröhler, D. (2014). Change management in the governance of schooling: The rise of experts, planners, and statistics in the early OECD. *Teachers College Record*, 116(9), 1–26. <https://doi.org/10.1177/016146811411600903>.
- Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21(1), 1–19.

# Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems

Susan R. Goldman, *University of Illinois Chicago*

Carol D. Lee, *Northwestern University*

## CONTENTS

INTRODUCTION .....	49
THE COMPLEXITIES OF HUMAN LEARNING AND DEVELOPMENT .....	51
Learning Is Participation in Cultural Practices .....	52
Constancies in Learning and Developmental Processes.....	53
Principles of Human Learning and Development .....	54
IMPLICATIONS FOR TEACHING, LEARNING, AND ASSESSMENT.....	55
Designing Learning Environments as Communities of Cultural Practices .....	56
Implications for Assessment Systems.....	59
REALIZING INSTRUCTION AND ASSESSMENT IN THE CONTEXT OF THE COMPLEXITY OF HUMAN LEARNING AND DEVELOPMENT .....	62
Case 1: Mathematics: Examining Everyday Repertoires of Practice as Linked to Disciplinary Learning .....	63
Case 2: Science: Relationships Among Discourse Registers.....	66
Case 3: Literacy: Problems of Figuration .....	69
Case 4: History: Designing for Historical Reading and Reasoning .....	71
Case 5: Recruiting Everyday Repertoires to Support Disciplinary Conceptual, Procedural, and Epistemological Knowledge in Tandem with Identity Development and Engagement: Cultural Modeling.....	73
Case 6: Building Teacher Professional Learning Communities as Central to Building Capacity for Learning, Teaching, and Assessments: Chèche Konnen .....	76
Case 7: District Level: Research Practice Partnership .....	78
Case 8: Systemic Work in Assessment: PISA.....	81
Cross-Case Analysis .....	83
CONCLUSION .....	85
REFERENCES .....	86



## INTRODUCTION

Advances in understanding the complexity and fundamentally cultural nature of human learning and development are foundational for efforts to achieve equity in educational systems at all levels: classroom instruction, curriculum, and assessment; district, state, and federal accountability assessment; and teacher preparation and learning in practice. In this chapter, we draw heavily on analyses and syntheses of human learning and development research from multiple perspectives (e.g., sociocultural, cognitive, epigenetic) and across multiple disciplines (e.g., neurosciences; cognitive, developmental, and social psychology; anthropology; learning sciences and education sciences). The conception we outline herein acknowledges how multiple dimensions of human development and learning (e.g., cognitive, social, emotional) are interdependent and culturally situated. Accordingly, all facets of educational systems need to focus on the whole child, as well as the multiple communities in which a child is situated.

The conception we outline asserts that learning unfolds through participation in the cultural practices of families and communities—including school classrooms, disciplinary communities (e.g., mathematics, science, social studies, literature, the arts), and out-of-school interest and affinity groups. It emphasizes the processes, not just the outcomes, of human learning. These processes encompass psychosocial dimensions of development (e.g., identity, resilience, mindset), including the emotional dimensions of such development, as well as the cognitive. Furthermore, this conception makes clear that diversity is a fundamental characteristic of the human species. It is thus essential to understand the diversity learners bring to formal and informal learning contexts, the pathways their learning takes, the support they need to make those journeys, and the outcomes of their learning and development.

We are not the first to argue for an expanded view of the goals of schooling if educational systems are to be able to prepare citizens for life in the 21st century. As we entered the 21st century, various sectors of society noted the transformation of the U.S. and global economy from one rooted in assembly line mass production to one that valued innovation, creativity, and relational rather than individual ways of working (21st Century Workforce Commission, 2000; Ananiadou & Claro, 2009; Binkley et al. 2012; Griffin et al., 2012; Voogt & Pareja Roblin, 2012). Dubbed by some the “knowledge society,” competencies deemed important for success entail problem solving, collaborative work, and flexible knowledge to support its use in new and novel situations. “Deep learning” emerged as the term to capture the contrast between learning to reproduce content and procedures in contrast to instruction that aimed for principled knowledge that allowed knowledge learned in one context to transfer and be useful in new situations (National Research Council, 2012; William and Flora Hewlett Foundation, 2013). There was also a push to expand the range of targeted competencies beyond cognition, including what the National Research Council (2012) referred to as intrapersonal and interpersonal competencies. The intrapersonal include intellectual regulatory, monitoring, and evaluative competencies and the interpersonal include collaborative, leadership, and communicative competencies. Fullan (2015) proposed a similar set of “deep learning” goals that he referred to as the 6 Cs: character education, citizenship, communication, critical thinking and problem solving, collaboration, and creativity and imagination.



Character education and citizenship, he argued, were essential to individuals' well-being and positive relationships with others (Fullan, 2015).

In a similar vein, the recent National Academy of Education report *Educating for Civic Reasoning and Discourse* argues that public education, in particular, plays an important role in preparing young people to engage in civic reasoning and discourse (Lee et al., 2021). The complex structure of democracy in the United States is designed with pathways to assist in navigating differences. This type of democratic republic requires both knowledge of the structures of governance and that citizens—broadly defined to include all who live in the country—embody dispositions to empathize with others, weigh multiple points of view and evidence, and value complexity over simplistic responses to complex problems. Thus, an expansive, equity-focused system of education and its assessments should address these dispositional goals as well.

In contrast to this expansive view of educational goals, traditional educational goals and assessments have been focused on achievement narrowly defined as cognitive skills, procedures, and an established canon of information (see Chapter 1 of this volume, “Reimagining Balanced Assessment Systems: An Introduction”). Progress through school has been measured by normative definitions of adequate progress—typically one year’s worth—along these cognitive dimensions. Whether through the National Assessment of Educational Progress (NAEP) or state assessments, there is a persistent history documenting disparate learning outcomes associated with race/ethnicity and class (de Brey et al., 2019). These assessment outcomes have been used to sort students, leading to widespread tracking and an absence of robust learning opportunities for racially and ethnically minoritized students, students from rural areas, and students whose families live in poverty (Legette, 2018; Lucas, 1999; Oakes, 2005; Tyson, 2011). In addition, outcomes from state and district assessments have been used for accountability, often resulting in even more restrictive teaching under the assumption that students with presumed lower-level skills cannot be taught complex conceptual knowledge. Balanced assessment systems—as defined in Chapter 1 of this volume, “Reimagining Balanced Assessment Systems: An Introduction”—require reframing the purposes of assessment from sorting and accountability to providing actionable information about student thinking and learning, information that supports teachers in meeting students where they are, and providing culturally responsive and sustaining learning opportunities (Armour-Thomas et al., 2019; Evans, 2021).

Enacting balanced assessment systems that are aligned with 21st-century conceptions of the goals of schooling requires reckoning with the complexity of human learning and development and acknowledging its fundamentally sociocultural and situative dimensions (Greeno & Middle School Mathematics through Applications Project Group, 1998; Nasir et al., 2021; Rogoff, 2003). Contemporary research on human learning and development calls for reconceptualizing assessment to reflect cultural, social, emotional, and cognitive dimensions (National Academies of Sciences, Engineering, and Medicine, 2018). Broadening and expanding what is to be learned and how it is learned requires changes in both what is assessed and the systems that do the assessing (Darling-Hammond & Conley, 2015). At the student level, assessments need to consider students' cultural repertoires, the ecological systems that support their learning inside and outside of school, and how these interact with assessments of the multidimensional components of learning (e.g., the 6 Cs [Fullan, 2015], the three sets of

competencies delineated by National Research Council, 2012; and the civic reasoning and discourse goals identified in Lee et al., 2021). At the classroom level, the what and how of instruction and the what and how of assessment should be both aligned and coherent. At the systems level, interpretations of assessment outcomes should account for key indicators that mediate outcomes—such as Opportunity to Learn (OTL; Marion, 2020)—and the availability of resources for teachers to create classrooms rich in OTL (e.g., time and material resources as well as structures for ongoing teacher learning).

In brief, we argue that robust balanced assessment systems will yield the most productive information if informed by a comprehensive understanding of the complexities of human learning and development. In this chapter, we first outline a basic contemporary understanding of these complexities, and then consider the implications for equitable classroom learning, instruction, and assessment.

### THE COMPLEXITIES OF HUMAN LEARNING AND DEVELOPMENT

The fundamental propositions regarding human learning and development reflected in this chapter are informed by a variety of disciplinary research, including developmental sciences, cognitive sciences, learning sciences, educational science, neurosciences, social psychology, and anthropology (Lee, 2017; Lee et al., 2020; Nasir et al., 2020; National Academies of Sciences, Engineering, and Medicine, 2018). The dispositions entailed in human learning are rooted in our evolution as a human species, including making sense of experiences, developing supportive relationships, “reading” others’ internal states, and feeling efficacious and safe (Cole, 2007; Quartz & Sejnowski, 2002; Tomasello, 2021). Thinking and learning are not solely cognitive activities—knowledge construction and organization involve motivational, affective, perceptual, and conceptual dimensions. The process of making sense of experiences involves recruiting prior knowledge as a resource for engaging new learning; understanding perceptions of the self along multiple dimensions; and analyzing perceptions of tasks, including the relevance of settings and relationships with others in those settings. These dimensions are in dynamic relationship with one another during learning.

Learning is anything but a passive process. Psychological, developmental, and neurosciences research have established that learners actively interact with people, other animals, objects, and physical environments (National Academies of Sciences, Engineering, and Medicine, 2018; National Research Council, 2000). Learners select what they attend to, as well as how they interact and with whom. These decisions are influenced by the learners’ perception of themselves along multiple dimensions, what they perceive as relevant knowledge, the emotional salience they attribute to their experiences, the resources they recruit from the historical moments of their life experiences, and the repertoires they employ from their participation in multiple cultural communities of practice (Spencer, 2006). Perceptions of the self include a sense of self-efficacy, motivation, and relevance. Indeed, there are well-established correlational relationships among academic resilience (e.g., dealing effectively with challenges, setbacks, adversity, and pressures in the academic setting), motivation, self-efficacy (confidence as a learner), and persistence (e.g., Martin & Marsh, 2006).

The correlational findings are supported by experimental behavioral and neurosciences studies that have established that the cognitive/knowledge and psychoso-

cial dimensions of learning intersect in significant ways (Osher et al., 2018). On the behavioral side, Dweck (2006) and Good and Dweck (2006) report that people's beliefs about the nature of intelligence impact their motivation to learn and their willingness to exert effort during learning tasks. Those who believe that intelligence is fixed and unalterable are less likely to persist and invest effort in academic tasks—especially if they are challenging—compared to those who believe that intelligence is malleable. Cognitive neuroscience provides evidence that thinking, feeling, and perceiving operate in dynamic relations with one another at the neural level. As learning occurs, regions of the brain associated with social and affective processes are activated, along with regions associated with cognitive processes and executive functioning (e.g., Damasio, 1995; Immordino-Yang & Damasio, 2007).

Neural pathways in the brain evolve as humans observe, imitate, interact with, and take their cues from those in their cultural, experiential, and interactional contexts. Although the brain is most malleable from infancy through adolescence, it retains its plasticity across the life course (Cantor et al., 2018). Contrary to conceptions that the brain is wired at birth and will never change, neural pathways and connections are responsive and transform throughout the lifespan. Understanding this reality contributes to growth mindsets that promote persistence—especially in the face of challenging tasks.

### **Learning Is Participation in Cultural Practices**

Learning is fundamentally social: humans interact with other humans and the cultural artifacts that human communities create across time (Rogoff, 2003). Cultural artifacts are manifestations of the routine cultural practices of a community, including their belief systems, systems of knowledge, routinized forms of social interaction, the tools for problem solving that they create (Cole, 1998), and the ways of using language that characterize participation in that community's practices (Gutiérrez & Rogoff, 2003; Rogoff, 2003). People participate in multiple communities of practice across multiple settings (Bronfenbrenner & Morris, 1998), beginning with the family and broadening to extended family groups and age- and interest-based communities (e.g., infant and toddler groups, sports clubs, school communities). Individuals begin as peripheral participants in these communities and gradually—through observation, imitation, and incrementally increasing their involvement in the community's routine practices—develop into fully participating members (e.g., Lave & Wenger, 1991). In moving from peripheral to full participation in these groups, learners adopt and adapt the group's discourse, norms, and values. Social relationships and attachments and perceptions of safety, self-efficacy, and relevance matter for development and goal setting (Bandura, 1993; Barron, 2006). In particular, perceptions of the self develop and unfold over the life course—including of the self as an individual, as a member of social groups, and as a member of cultural communities characterized by routine cultural practices and belief systems that evolve and have longevity over time and space.

Thus, participation in cultural practices and social interactions is essential to human development from the moment of birth in all areas of development: Learning language, learning to infer the internal states of others—human and animate, learning to walk and learning to manipulate objects (Gopnik et al., 1999; Kuhl & Meltzoff, 1996; Lee et

al., 2020; Meltzoff, 1988, 2013). In the case of language, at birth, infants can hear all the sounds and phonemes of all human languages (Meltzoff et al., 2009). However, through imitating, hearing, testing, and reproducing the sounds of the language or languages around them, infants' neural networks undergo a pruning process to hone in on the functionalities of the language or languages of their social environment (Kuhl, 2011; Kuhl et al., 2014; Meltzoff & Kuhl, 2016). Language development also illustrates that human learning is an outgrowth of biological processes, taken up as people engage in the routine cultural practices of a range of social groups, beginning with the family and extending outward to peers, affinity groups, community-based groups, etc.

The diversity of cultural practices in which an individual participates necessarily gives rise to variation in pathways, processes, and timing of what is taken up by whom and under what conditions. That is, while there are fundamental tasks to be accomplished at different stages of the life course—particularly from infancy through adolescence—how these tasks are learned and what social and individual functions they play are diverse, influenced by the communities of routine cultural practices in which the individual develops (Rogoff & Chavajay, 1995). Diversity is also crucial to adaptability and survival. For example, diversity in the gene pool increases the resilience of a species and the likelihood of survival in the face of extreme threats (Booy et al., 2000). Thus, diversity in pathways of development is both normal and essential for the evolution of the species.

### **Constancies in Learning and Developmental Processes**

At the same time that we can expect diversity in how individuals accomplish fundamental tasks at different life stages, developmental theorists identify several constancies of development and learning. One such constant is the homeostatic principle: systems seek to maintain balance. Jean Piaget referred to this as equilibration: organisms strive to achieve a balance between the new (accommodation) and the old (assimilation) (Ginsburg & Opper, 1988). Humans strive to balance the degree to which they change in response to new experiences and social interactions against the degree to which they fit new ideas or experiences into their existing conceptions of the physical, cognitive, and social worlds. In the social realm, Heider (1958) argued that humans seek balance in relations by choosing new groups as we change or changing our beliefs and thinking to fit the groups we are in.

Continuity and change as constant oppositional forces in learning and development also play a central role in Lev Vygotsky's sociocultural theory (van der Veer, 2014). For Vygotsky, development and learning could only be understood by considering both what is known and what is yet to be learned, with the latter reflecting the process of learning. He proposed that accounting for development and learning required the consideration of what children could do on their own (what is known) and what they could do in collaboration with adults, labeling this the *zone of proximal development* (ZPD) (Zaretskii, 2009). What the learner can do on their own reflects what has developed; what they can do with the assistance and guidance of a more knowledgeable other—typically an adult but also possibly more knowledgeable peers—is the process of learning. Providing effective guidance depends on assessing what the learner can do on their own in conjunction with an understanding of what constitutes “next steps” in



that individual's learning process. In other words, providing effective guidance must consider where the learner is as well as where they are going—what processes have completed their cycle of development as well as what processes are still in development. Zaretskii (2009) also pointed out that while Vygotsky died before setting out the pedagogical implications of the ZPD, concepts such as diagnostic assessment and dynamic assessment (Feuerstein, 1979; Feuerstein et al., 2002) are outgrowths of Vygotsky's broadening the conception of development to include not just what has developed but the learning process that creates future developmental outcomes.

Finally, as humans participate in cultural practices and what they know about the world and themselves grows, information becomes more differentiated, creating a need for organizing systems (Werner, 1957). An apt example of differentiation comes from Nelson's (1973) account of word learning. She proposed that during early word learning, labels are applied in accordance with the functions of objects. That is, anything that is round, rollable, and/or throwable may initially be labeled *ball*. This concept becomes differentiated as toddlers interact with perceptually round objects that vary in size, texture, squeezability, bounceability, and so on. At some point, toddlers distinguish between round things and create functionally relevant distinctions (e.g., balls that bounce, balls that we eat such as apples or oranges). With respect to content taught in school, similar reorganizations should be expected as students acquire content through their experiences in communities of classroom practice. Organizational systems must be adaptive and allow individuals to use what they know to function flexibly in response to changing environmental and contextual conditions.

### **Principles of Human Learning and Development**

Box 3-1 summarizes the foundational principles of human learning and development we have discussed. These principles are at the core of efforts to develop balanced assessment systems and practices as defined in this volume (see Chapter 1 of this volume, "Reimagining Balanced Assessment Systems: An Introduction"): systems that are centered and designed to function at the classroom level to provide teachers and students with feedback that guides instruction and supports students with appropriate learning opportunities. Developing such systems requires expanding and differentiating among the existing purposes and functions of assessment (e.g., Bennett, 2011; Penuel & Shepard, 2016; Shepard, 2019). That is, balanced assessment systems must expand beyond summative assessments *of* learning, often used to sort students, to focus on assessment *for* and *as* learning (Black & Wiliam, 2009; Wiliam, 2011). Assessment *for* learning provides information about where students are relative to where they are headed, thus informing ongoing instructional planning to support further learning (e.g., Bennett, 2011; Penuel & Shepard, 2016; Shepard, 2019). Assessment *as* learning reflects the inherently social and cultural nature of learning, a principle fundamental to our argument in this section. Assessment *as* learning focuses on the process of learning as it is happening and is visible to participants in the learning situation (Bennett, 2011; Penuel & Shepard, 2016, Shepard, 2019). All three types of assessments should attend to the range of prior knowledge, dispositions, and belief systems that learners bring to new opportunities to learn. Admittedly, this is a complex mandate to achieve but it is critical to creating anti-racist and equitable educational systems.

**BOX 3-1**  
**Foundational Principles of Human Learning and Development**

- Learning entails dialogic relations among thinking, emotional salience attributed to experience, and perceptions of the self along multiple dimensions
  - Thinking and the role of prior knowledge in new learning
    - Conceptual knowledge
    - Procedural knowledge
    - Epistemology
  - Emotional salience
    - Perceptions of safety and self-efficacy (e.g., growth mindset)
    - Perceptions of relevance
  - Perceptions of the self
    - As an individual
    - As a member of cultural communities of practice (family; social networks; interest networks; and institutional settings such as schools, community organizations, age cohorts, etc.)
- Relationships matter
- Participation in routine cultural practices within and across settings, within and across time
  - Affordances of artifacts, belief systems, and practices in cultural communities
  - Relationships across different cultural communities of practice
  - Where learners are in the life course
- Learning is malleable across the life course
- Diversity in developmental pathways is essential for the human species

**IMPLICATIONS FOR TEACHING, LEARNING, AND ASSESSMENT**

The foundational principles of learning and development we have identified imply several goals for designing robust and equitable learning environments, several of which are included in Box 3-2. New learning needs to build on what learners already know, based on their experiences as well as the language and discourse practices with which they are familiar. At the same time, more knowledgeable others (e.g., adults, peers) with whom learners interact introduce new ways of thinking, reasoning, and using language in ways that extend but connect to what learners already know and do. Making thinking and knowledge construction processes visible and objects of reflection makes the processes learners are engaged in concrete, helping to bridge the new and old, creating a means of balancing assimilation and accommodation processes. Periodically, learners will shift how they organize what they know to reflect differences and similarities that become noticeable as they accumulate more information through interaction with others and objects in their environments. Such reorganization can be expected to have functional value for the learner, making routine tasks and cultural practices more efficient and effective. Assessment that targets processes of reorganization has the potential to contribute new insights to the diversity of developmental trajectories that characterize human learning.

The design goals specified in Box 3-2 should guide the goals, purposes, and motivations for making decisions about curriculum (what content, principles, and perspec-



**BOX 3-2**  
**Goals for Designing Robust and Equitable Learning Environments**

- Connect students' prior knowledge and experiences across multiple domains to new learning targets
- Build nurturing relationships
- Make reasoning processes public
- Focus on rich conceptual knowledge and the practices by which it is generated
- Support students as they engage in inquiry, knowledge-building, and reorganization
- Support students in seeing and understanding the relevance of learning targets to students' perceptions of their needs
- Support and position students as self-efficacious

tives will be taught), instruction, and assessment practices. Goal setting and purposeful learning arise out of perceptions of the epistemic goals and relevance of the task(s) (i.e., why am I doing this?), self-efficacy about the task(s), and relationships with others in the setting (Eccles & Wigfield, 2002). These perceptions undergird motivation to engage, persist, and achieve the end goal. When learners have little understanding of the purposes of the work they are doing and do not perceive the relevance of the tasks they are asked to complete, they are likely to have little to no motivation and exert minimal. Epistemic purposes that contribute to robust and equitable learning include valuing complexity and inquiry. These purposes contrast with memorizing facts and procedures “for the test.” Teaching and assessments need to pay greater attention to differences in learners’ perceptions of epistemic purpose and relevance and work toward those that foster engagement in inquiry and grappling with complexity.

In addition, realizing the design goals shown in Box 3-2 requires embracing a vision of instruction and assessment as participation in communities of cultural practices. The norms, discourse, values, and goals inherent in cultural practices are negotiated and re-negotiated by the members of the community, making a shared sense of agency, authority, and ownership possible (Gee, 1992; Lave & Wenger, 1991). Increased attention to where and with whom the agency and authority for making these decisions lies is fundamental to achieving anti-racist equitable educational systems. Deliberations around such decisions need to involve multiple stakeholders, including teachers, parents, students, community stakeholders, and governmental authorities. Ideally, such deliberations are informed by commitments to democratic principles and equitable goals.

**Designing Learning Environments as Communities of Cultural Practices**

Learning environments designed to create communities of cultural practices are not only consistent with the complexity of human learning and development but can also be a powerful means of supporting active, agentive learning in educational settings (Lee, 2010). Classroom communities of practice engage students in the active construction of knowledge, asking them to wrestle with conundrums that arise in their inquiries and to

work independently and collaboratively to make sense of often conflicting information, perspectives, and values. The goals of school-based communities of cultural practices can include both knowledge construction in the moment and individual and collective development in the future.

The sensemaking processes in which students engage in classrooms designed as communities of practice are developmentally appropriate forms of the knowledge generation processes and practices engaged in by members of professional disciplinary communities. Rather than simply learning facts and procedures, classroom communities of practice engage students in doing intellectual work that approximates professional disciplinary practice. Knowledge is generated or constructed through interactions with others, material resources, and goals (e.g., Vygotsky & Cole, 1978). As such, these classrooms instantiate a form of apprenticeship in which more knowledgeable others (e.g., teachers, mentors, tutors, and peers) make knowledge generation processes (cognitive, intrapersonal, interpersonal) and the results of those processes (solutions to problems, theories and revisions of theories, tools) visible (Collins et al., 1989). These interactions and observations become the basis of internalized knowledge representations and include cognitive (memory, perceptual, reasoning process), social, and affective dimensions. In turn, what has been internalized shapes how learners experience and observe subsequent interactions—how and what learners think and feel arises from complex interactions that reflect learners' cultural and contextual circumstances. The individual and their community are changing and evolving together through their joint participation (Rogoff, 1997). Learning is then defined, in part, as the transformation of an individual's participation in valued social and cultural activities. Such learning can also involve transformations of what social and cultural practices are valued. Which processes an individual engages in can involve emotional, motivational, and relational aspects of self—not just knowing (Holland & Lave, 2009).

Content domains or disciplines can be conceptualized as communities of practice wherein the members negotiate and re-negotiate the norms, conventions, and criteria for proposing, arguing for, establishing, and evaluating knowledge claims and the arguments put forth to support them (Lave, 2012; Lave & Wenger, 1991). Criteria are established for what constitutes valid and reliable inquiry practices, including patterns of logic and reasoning for connecting evidence to the claims it is intended to support (Toulmin et al., 1984). Evidence that does not meet these criteria compromises whatever claims are being made on the basis of that evidence. Members of a disciplinary community also share common epistemic commitments to the aims, goals, and purposes of argument and knowledge generation within their discipline (e.g., explanation, evaluating alternatives, proposing policies) and the representational forms used to communicate their ideas with one another (e.g., Bazerman, 1985; Goldman et al., 2016; Shanahan et al., 2011). It is important to note that disciplinary communities often invite contestation and diversity of aims and goals (e.g., Knorr Cetina, 1999). Such dispositions are often reflected in contested power relationships manifesting from the history of who has contributed to specific norms, modes of reasoning, and forms of representation. One example includes recent attention to Indigenous knowledge systems as scientific conceptions of the relationships between humans and the rest of the natural world (Bang & Marin, 2015). At the same time, it is important to recognize that disciplines evolve. As Kuhn (2012) notes, the evolutionary histories of a variety of disciplines are

replete with consequential shifts in assumptions about what counts as evidence and epistemic goals. For example, Osborne et al. (2003) have argued that students need to study the history of science to understand how epistemic shifts unfold and why, leading to a deeper understanding and appreciation of science itself. As teachers navigate the integration of students' cultural repertoires into content area instruction, knowledge of these epistemic negotiations is important. These epistemic negotiations are also important for those engaged in assessments (e.g., diagnostic, formative, summative) so that they might bring such breadth of knowledge to assessment design. We illustrate the relationships between everyday cultural repertoires and disciplinary knowledge in the final section of this chapter.

As such, instruction and assessment based on the goal of engaging learners in developmentally appropriate forms of the cultural practices of the discipline apprentice learners into becoming full participants in *classroom* disciplinary communities replete with the cultural practices of knowledge generation, including multiple ways of participating in disciplinary inquiry. In such classrooms and learning situations more broadly, learners build an understanding of disciplinary concepts, principles, processes, representational forms, discourse genres, and conventions of language use through inquiry processes and problem solving.

One issue for consideration in the context of anti-racist and equitable instruction and assessment design is how decisions are made regarding what is developmentally appropriate and what constitutes acceptable diversity in conceptions of the disciplines. Do those decisions reside at local, national, or federal levels? If manifest in standards, what latitude is there for adaptation at the local level to address learners' diversity in experiences, language, goals, and values?

### *Challenges of Design and Implementation*

There are several challenges inherent in designing and implementing communities of disciplinary practice in classrooms. First, it is important to recognize the complexities of what is presumed a disciplinary community of practice, and by whom. Some disciplines—as in the study of literature or history—value debate on the logic of argumentation and place high value on relations between claims and evidence. Considerations of what it means to recruit repertoires of knowledge, practice, and discourse from students rooted in diverse cultural communities requires revisiting what constitutes disciplinary practice. Indeed, this is not a challenge isolated to students from diverse cultural communities because, typically, few students enter classrooms having experienced the discourse and formal practices of disciplines. Such considerations may include theorizing what are the conceptual and discursive relationships between everyday cultural practices and an academic discipline; rethinking the historical evolution of knowledge within a given discipline; and/or examining the diverse forms of reasoning and representations of concepts that may be captured across different cultural communities. Indeed, how federal and/or state standards are specified as “developmentally appropriate” is bound up in issues of who decides what is valued and the latitude afforded for adaptations at various levels of the educational system (e.g., specific districts, schools, teachers, students), adaptations that value and respect the specific implementation context. The point here is that connecting the repertoires that students bring into the

classroom with what may be identified as formal disciplinary knowledge is a complex undertaking. Due to this complexity, thinking about how to design instruction and assessments that make these connections visible requires systemic support.

Effective learning spaces are safe, efficacious, and driven by attachments to other people. Such spaces value and build on cognitive, social, and cultural resources that individuals bring to learning situations—whether those situations are formal schooling, virtual, or community spaces. Dispositions toward learning, identity as learners, and resilience in the face of adversity are important orientations to formal and informal instructional situations. For example, instructional literacy programs that build on Indigenous Hawai’ian narrative participation structures, African American narrative structures, and rapping all welcomed and valued linguistic repertoires not typically sanctioned in formal school settings (Au, 2013; Champion, 1997, 2003; Emdin, 2013; Pinkard, 1999). In these situations, students were invited into the conversation in ways that made safe spaces for participating in the linguistic practices of the school curriculum. Later in this chapter, we offer case studies of teaching, learning, and assessment designs that push the boundaries of traditional conceptions of particular disciplinary communities of practice.

The multidimensional nature of human learning implies that classrooms organized as disciplinary communities of practice need to attend to the interconnected social, emotional, cultural, and cognitive facets of learning and development. This attention includes critical examination of what is assumed to constitute disciplinary communities of practice. Conceptualizing disciplinary communities of practice in the context of schooling involves understanding distinctions between expectations in professional disciplinary communities and the developmental demands of disciplinary learning in schools, as well as examining the historical and political influences on how a given discipline is represented institutionally. Creating and valuing multiple pathways and trajectories is fundamental to achieving equitable instruction that is safe, supportive, and efficacious across a broad spectrum of learners.

### **Implications for Assessment Systems**

Balanced assessment system design is a critical aspect of achieving equitable learning environments. As cogently argued by Shepard (2021) and Shepard et al. (2018), assessments that inform learning must be closely aligned to where and how learning is happening, as well as how that learning and its assessment are supported. In the context of schooling, that means assessment must be closely aligned with instructional practices, processes, and outcomes. Thus, assessment must be reconceptualized, designed hand in hand with instruction, and both assessment and instruction need to attend to knowledge as well as the social and affective dimensions of learning. In the following, we briefly discuss the limitations of currently available instructional and assessment materials to show why a fundamental and thorough reconceptualization of both is necessary to address the cultural foundations of learning.

### *Limitations of Existing Commercial Products*

Most commercial curricula include instructional practices embedded in materials and tasks, as well as resources for assessing student learning. However, there are few examples of commercial instructional materials, including curricula and assessments, that embody the principles of learning and development discussed in this chapter. Even diagnostic assessments fall short, even though they are typically closest to instruction on the ground and are intended as assessments *for* learning. Specifically, diagnostic assessments often provide a picture of only what a student has already mastered and can do independently (assessment *of* learning), ignoring the exploration of what the student can do with support or in collaboration with peers, tutors, or teachers (assessment *as* learning). The concept of learning progressions, however, can contribute to teaching and the design of assessments that consider what students can do with support (Duncan & Hmelo-Silver, 2009; Duschl et al., 2011). Research on learning progressions, most focusing on mathematics and science, identifies conceptual relationships between lower-level skills and higher-order reasoning. Typical diagnostic assessments do not capture many important aspects of the learning process because they are often constrained to a subset of facets of the knowledge dimension. There are, for example, no easily available commercial diagnostic assessments of epistemological orientations, perceptions of self or tasks, recruitment of cultural repertoires into sense-making and problem solving, or engagement with disciplinary practices. Understanding what students can do with support requires understanding the relationships between what students already know and the demands of the “next level” of tasks. Furthermore, pathways of support must go beyond the cognitive, to attend to epistemology, understandings of the self and the task, and of relationships among those engaged in the learning process.

Existing assessments are also severely limited with respect to the aspects of disciplinary learning that are assessed, especially for social studies, science, and the arts. In reading comprehension—whether diagnostic, formative, or summative in function—few if any assessments provide insight into how students reason with texts. Furthermore, although knowledge of academic language is essential to reading, writing, and discussion within and across academic disciplines, it is not widely assessed. This is so despite the existence of a useful assessment of academic language, namely *Core Academic Language Skills* (Uccelli et al., 2015). Assessments of writing tend to focus on mastery of rhetorical structures with a lesser emphasis on content, logic, and reasoning. Writing assessments also rarely if ever consider the functions of language variation as rhetorical tools. Finally, although we have emphasized the importance of perceptions of safety and belonging in learning, attention to social-emotional learning is typically siloed and not integrated into content area instruction and assessment.

### *Designing to Address the Cultural Foundations of Learning*

In brief, we argue that what is needed is a fundamental reconceptualization of how learning-centered assessment addresses the cultural foundations of learning. Thinking, feeling, and perceptions are intertwined, and it is these relationships that fuel learning. Accordingly, instruction and assessment must address these relationships. Both instruction and assessment need to encompass the breadth of knowledge, problem



solving, and epistemological orientations within content domains, as well as the history of knowledge construction that accurately captures students' contributions across time and space. In addition, instruction and assessment need to encompass conceptions of possible and expansive futures and the kind of supports needed to propel students to realize these futures. Undertaking such a reconceptualization is a prerequisite for ensuring that assessment plays a pivotal role in providing actionable information regarding the full range of competencies that inform and explain learning. Critically, such a reconceptualization needs to address issues of equity for communities that face persistent structural challenges related to race, ethnicity, gender, socioeconomic status, perceptions of ability, or language variation.

Learning-centered assessments need to provide insights and feedback about the following issues:

- Opportunity to learn as reflected in, for example, teacher quality, curriculum quality, and access to expansive learning
- What knowledge and reasoning is valued and worth assessing as well what counts as legitimate displays of that knowledge and reasoning
- The consequences of assessments for students' identity with respect to the disciplines and their future orientations to disciplinary pathways

Interrogating these issues requires rethinking ongoing assumptions about what constitutes knowledge in the disciplines. For example, research on Indigenous knowledge systems about the natural world indicate an epistemological orientation that centers culture–nature relations as lived (Bang & Marin, 2015). This Indigenous orientation is typically not considered in assessments of scientific reasoning even for students from Indigenous communities who have been socialized with this orientation. Similarly, research in the field of ethnomathematics documents diverse practices that entail mathematical reasoning but that do not look like the standard practices that are valued, instructed, and assessed in the majority of U.S. schools (Ascher, 1991). This is also the case for language instruction and assessment in that both ignore research showing the affordances and functionalities of language diversity and variation from so-called Standard English (Smitherman, 1995). Thinking broadly about assessment regarding issues of equity also requires a reconsideration of the concepts of group membership, particularly regarding the constructs of race, ethnicity, and gender. It is important to avoid what Gutiérrez and Rogoff (2003) call the “box problem,” or assuming homogeneity within cultural communities.

Although assessments are typically thought of as tests given at different time scales during instruction, it is also important to consider the informal assessments that teachers make in the moment—“on the fly”—as instruction unfolds. These serve important formative purposes—assessment *for* and *as* learning, depending on the way the assessment unfolds and how the information is used. For example, looking at student work while it is being produced may provide the teacher with valuable information about what a student does and does not understand, and thus what instructional “next steps” would be useful. Often, such embodied assessments unfold in class discussions. This kind of informal assessment during instruction is important for moment-to-moment responding as well as day-to-day planning. Such assessment requires that

teachers have deep pedagogical content knowledge and expansive understanding of child and adolescent development (in the K–12) classroom to understand what student responses, assertions, and representations convey, and the implications for where students are with respect to instructional goals (Goldman & Snow, 2015). For example, Magdalene Lampert (1990, 2003)—a math education researcher, grade 5 math teacher, and university faculty member—illustrated in detail how she made assessments of student learning during instruction, the knowledge base she drew from to respond in the moment, and how to adapt her plans for subsequent days. Similar work has been carried out by Ball (Ball & Cohen, 1999), who recruited what she learned from her own teaching to inform professional development for university students studying to become teachers. There is also an expanding literature on the use of these types of informal assessments to inform responsive instruction in other disciplines (e.g., Elby et al., 2014; Jaber et al., 2022; Michalchik & Gallagher, 2010; Ruiz-Primo & Furtak, 2007).

Box 3-3 summarizes the foregoing discussion of the characteristics and design considerations of balanced assessment systems that are aligned with equitable learning environments that reflect the complexity of human learning and development.

### **REALIZING INSTRUCTION AND ASSESSMENT IN THE CONTEXT OF THE COMPLEXITY OF HUMAN LEARNING AND DEVELOPMENT**

The breadth of what influences learning and its implications for assessment can be overwhelmingly complex. In the following sections, we offer cases of instruction and assessment that embody the principles discussed in this chapter. The cases connect teaching, learning, and assessment to students' knowledge and repertoires developed through their participation in everyday, routine cultural practices. They articulate the breadth of what disciplinary knowledge entails and connect teaching and assessments in ways that reveal the breadth of relevant student knowledge and dispositions. They

#### **BOX 3-3**

##### **Characteristics and Design Considerations for Equitable and Learning-Centered Assessment Systems**

- Vertical coherence among classroom practices, class/school level assessments, and district/state assessments
- Provision of actionable data on opportunity to learn rather than accountability
- Addresses cognitive processes and practices of reasoning
- Addresses perceptions of:
  - self-efficacy
  - mindset
  - motivation
  - relevance
- Examines relations between multiple sources of prior knowledge and targets of new learning
- Makes students' reasoning and inquiry processes visible
- Addresses classroom climate, school culture, and district policies
- Includes learners and teachers in the design process

take place in real classrooms and involve real collaborations among teachers and researchers.

Cases 1 through 4 illustrate instruction and assessment that recruit students' everyday repertoires and connect them to learning goals in the disciplines of mathematics, science, literacy, and history. Each case identifies unique opportunities to learn the nuances of what students can do that are not typically considered in traditional curricula or assessments. Each case took place in a real classroom, was developed through collaborations between practitioners and researchers, pushing the boundaries of what it means to show competence in disciplinary problem solving. Each case illustrates the kinds of reflection required to consider relationships between students' everyday experiences—situated in cultural communities of practice—and the academic disciplines taught in schools. Case 5 demonstrates how building on these illustrations of scaffolding everyday knowledge to teach and assess disciplinary knowledge can provide possibilities for supporting and assessing outcomes beyond siloed cognitive knowledge. The final three cases speak to issues at the systems level. Case 6 illustrates how teachers' pedagogical reasoning can be supported through professional learning communities and how this reasoning can, in turn, support ambitious and equitable instructional practices. Cases 7 and 8 illustrate how the many dimensions of human learning and development can be taken up at the district level—Case 7—and the level of broader systemic assessments—Case 8.

### **Case 1: Mathematics: Examining Everyday Repertoires of Practice as Linked to Disciplinary Learning**

Case 1 demonstrates how careful observations of children in a particular community engaging in purchasing practices outside of school can shed light on the complexities of their understanding of base ten computation. A collaboration between researcher Edd Taylor and a practitioner, this work led to the design of a classroom-based assessment that situated problems in the context of the children's everyday purchasing practices (Taylor, 2009). The work provided a window into more nuanced understandings of children's computational skills than the traditional measures that had been used.

Taylor examined and documented how a population of African American children, aged 4–10, from a low-income community, engaged in computational reasoning when they purchased goods (e.g., candy, toys) from a local store. Taylor (2009) made the following observations:

Observation 1: A girl about the age of 8 collects and replaces different combinations of lollipops and small candies while looking at change in her hand. After a few moments of collecting higher and lower values of candy she walks up to the clerk and asks, "Can I owe you 20 cent?" Having correctly determined the amount she needed, and with the owner's consent, she places her change on the counter and exits the store with her purchase.

Observation 2: In a separate corner of the store a third-grade boy fingers through about \$8 in cash halfway pulled out of his front pocket. He negotiates with two of his classmates about how much he can loan them and still be able to buy all the items he has collected for himself.

Observation 3: A child places a quarter on the counter and grabs four pieces of “nickel-candy.” The clerk tells the child, “Grab another one,” referring to the nickel-candy. The child grabs a fifth piece of candy and departs. (p. 374)

These observations documented complex computational reasoning in an ecology of learning beyond the classroom. Children’s reasoning provided evidence of their understanding of different denominations of U.S. currency as artifacts (e.g., half-dollars, quarters, dimes, nickels, pennies), their purchasing motivations, and the supports provided by the store clerk and their peers. Having documented this system of purchasing, Taylor analyzed the types of mathematical problems present in each phase of this system and their relevance for teaching computational reasoning in the primary grades. Table 3-1 indicates the different phases of making a purchase and the mathematics involved. Taylor also analyzed supports available in the learning ecology of the store, as represented in Figure 3-1.

Based on these analyses of the computational reasoning and supports embedded in one of this community’s everyday cultural practices, Taylor worked with a teacher to create a mock store in the classroom—complete with a variety of items to be sold, the currencies that could be used to purchase the various items, and the computational reasoning involved in purchasing each of the items—as summarized in Table 3-2. As a designed artifact, the mock store afforded opportunities for the students to recruit their everyday practices in the classroom. This provided opportunities for the teacher and Taylor to examine and assess the children’s computational reasoning. The mock store provided an opportunity for informal assessment, situated in a classroom and designed to provide a window into both the computational strategies children use and their conceptual understandings of the currency artifacts.

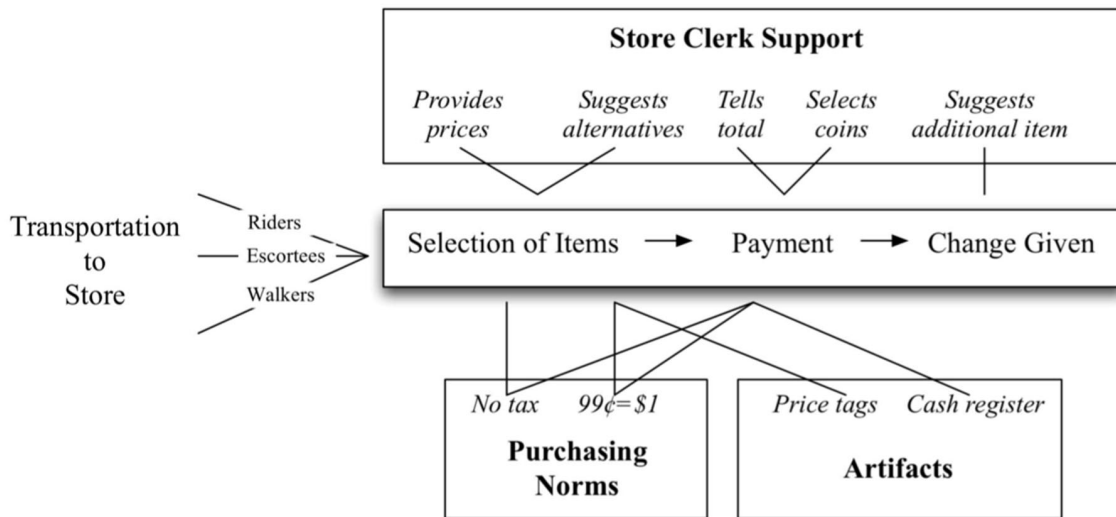
The design of this informal assessment provided opportunities for the teacher and Taylor to examine the students’ reasoning strategies, as illustrated in Table 3-3.

The rubric identified in Table 3-3 provided the teacher with detailed information about students’ reasoning processes—not simply outcomes of problem solving. Access to students’ reasoning enables teachers to plan subsequent instructional moves intended to move students’ thinking toward more successful strategies and reasoning processes.

**TABLE 3-1** Types of Mathematical Problems Children Engage During Phases of the Purchasing Practice

Phase	Mathematics Encountered
Selection	<ul style="list-style-type: none"> <li>• Reading of notational representations of prices (price tags)</li> <li>• Comprehension of number words spoken by the clerk</li> </ul>
Payment	<ul style="list-style-type: none"> <li>• Coin recognition and knowledge of coin value and equivalence</li> <li>• Addition of coin and bill values</li> <li>• Addition of item prices</li> <li>• Subtraction of total cost of items and amount money on hand</li> <li>• Equivalence relations of bills and coins</li> </ul>
Change	<ul style="list-style-type: none"> <li>• Estimation of change expected</li> <li>• Calculation of expected change</li> </ul>

SOURCE: Taylor (2009). Reprinted with permission.



**FIGURE 3-1** Influence of types of support at particular phases of purchasing.  
 SOURCE: Taylor (2009). Reprinted with permission.

**TABLE 3-2** Mock Store Shopping Lists, Purchase Totals, Mathematics Considerations, and Currency Available

List	List Items	Total of Items	Currency	Mathematics
A	1 bag of chips 1 lollipop	\$1.24	Quarters only	More than one dollar, coordinates dollars and cents, few items
B	1 box of cookies 2 pieces of taffy	\$1.20	All currency	More than one dollar, coordinates dollars and cents, few items
C	1 box of cookies 2 lollipops	\$1.50	All currency	More than one dollar, coordinates dollars and cents, more items
D	3 lollipops	\$0.75	All currency	Less than one dollar, more items
E	1 lollipop 2 pieces of taffy 1 piece of gum	\$0.50	All currency	Less than one dollar, more items

SOURCE: Taylor (2009). Reprinted with permission.



**TABLE 3-3** Definitions and Examples of Students' Mock Store Strategies

Strategy	Definition	Example
<b>More Successful</b>		
Total	Student determines the total amount needed through mental calculation and presents coins/bills together for payment.	Student grabs two pieces of taffy (10 cents each) and one piece of gum (5 cents). Pauses, thinks, then places one quarter on the table for payment.
One-to-one	Student matches each bill or coin to an item worth that amount.	Student places one dime next to the taffy, one nickel next to the gum, and one quarter next to the lollipop.
<b>Less Successful</b>		
Dollar-as-one	Student considers the value of cents and dollars as the same. When counting cents, child counts dollars as if they were one cent.	Student presents one quarter and one dollar as 26 cents.
Coin-as-ones	Regardless of the value of the coin, student counts the coin as being worth one or one cent.	Student counts collection of three nickels and three quarters as "one, two, three, four, five, six cents."
One-for-all	Students presents one coin for a multiple number of same-priced items.	Student presents one quarter to pay for three lollipops that cost 25 cents each.
Idiosyncratic	Student appears to use a strategy but one that does not follow any known logical pattern.	Student grabs a box of cookies and places a one-dollar bill on the table. The child then grabs two lollipops and places a random handful of change and calls it "three cents."
No strategy	Student appears to have no strategy because he or she is unable to provide payment or reports guessing.	"I guessed."
Unknown	The category could not be determined due to inadequacy of notes or audiotaping.	Not applicable

SOURCE: Taylor (2009). Reprinted with permission.

### Case 2: Science: Relationships Among Discourse Registers

Learners' understanding of science concepts may be underestimated if the language and context of the assessment are not aligned with experiences in which students understand the concepts, even if they express their understanding in non-technical terms. Relationships between everyday language and technical knowledge of disciplines—especially in science—are significant tasks to master. Case 2 spotlights Bryan A. Brown's research to illustrate the complexities of wrestling with these discourse relationships and how attention to language dimensions can be a useful focus for both teaching and assessment.

In one study, Brown and Kloser (2009) examined implicit understandings of physics concepts among high school baseball players and how these understandings mapped to formal physics principles. An ethnically diverse group of high school baseball players (11 African American, 7 Hispanic American, 7 Caucasian American, and 2 Asian American) were interviewed about their understanding of why a curveball moves as it does. The situation in question is illustrated in Figure 3-2 and shows the forces that influence the direction and speed of a curveball in the baseball context.



FIGURE 3-2 Causal factors for a curveball.  
SOURCE: Brown & Kloser (2009). Reprinted with permission.

The researchers were interested in what the adolescents understood, how their understandings mapped to the formal physics principles impacting the movement of the curveball, and the relationships between the language of baseball and the formal language of physics. Brown and Kloser (2009) framed these relationships as issues of conceptual continuity, consisting of cognitive and linguistic dimensions. Cognitive continuities focus on conceptual relations between everyday and disciplinary concepts (i.e., similarities and differences between students' conceptual understanding of velocity and speed in the informal baseball context compared to the formal high school physics curriculum). Language continuities refer to relations between everyday discourse and disciplinary discourse. Brown and Kloser (2009) argued that both sources of continuity and discontinuity between everyday and disciplinary practices needed to be addressed.

To address these relationships, an interview protocol was designed to ask questions in both baseball and physics genres, providing access to students' conceptual understanding across these two discursive contexts. For example, one form of a question used colloquial baseball language: "Please describe how the seams play a part in how the ball moves through the air when the pitcher attempts a curveball." This question was followed with a more canonical question using technical terms of physics: "Please describe how the seams affect the drag, velocity, and air pressure that affect the ball when a pitcher attempts a curveball" (Brown & Kloser, 2009, p. 879).

The researchers classified the students' responses into four categories: everyday discourse, baseball discourse, science discourse, and hybrid discourse. Table 3-4 defines each category and provides examples from student interviews.

An interesting finding from this study was that a formal, traditional multiple-choice science assessment administered before and after the baseball season showed no evidence that student performance had improved. However, in their responses to the interview protocol students' explanations of what forces impacted the curve and speed of the ball did show shifts in the discourse genres present, toward the inclusion

**TABLE 3-4** Modes of Discourse

Code Name	Code Description	Example
Everyday discourse	Instances where the player's descriptions of why curveballs curve involves the use of everyday (non-scientific/non-baseball) talk that is associated with baseball	Yeah, 'cause the—'cause when you throw the ball, the air is gonna hit the seams, so I guess that's the main point of making the curve ball
Baseball discourse	Instances where the player's description of why curveballs curve involves the use of genre-specific talk that is associated with baseball	If you throw a curve ball, the seams cutting through the air, it's gonna cut down
Science discourse	Instances where the player's description of why curveballs curve involves the use of science terms to explain why a curve baseball curves	So I guess probably the top one's high pressure and the bottom one's low and it's pushing it down so that it looks like it's curving
Hybrid discourse	Instances of talk where students explain science concepts using both blended versions of either science and baseball words or science and everyday terms associated with baseball to explain phenomenon	It doesn't break at all. I mean, it hangs—it actually didn't break because it had enough spin like a front spin on it so it would drop

SOURCE: Adapted from Brown & Kloser (2009). Reprinted with permission.

of more hybrid and science discourse postseason. Students were able to communicate their conceptual understanding using baseball registers, scientific registers, and hybrids of both. Based on their analyses, Brown and Kloser argued for the importance of understanding students' everyday practices that embody links to disciplinary concepts, and the importance of examining relationships between how such understandings are communicated in both everyday and disciplinary contexts:

We argue that viewing students' science understanding through two modes of conceptual continuity: (a) conceptual continuity as cognitive and (b) conceptual continuity as linguistic provides descriptive evidence of how students' understanding exists at varying levels of continuity with science ideas. These continuities are critical in enabling students to use their native ways of understanding the world in meaningful ways. (Brown & Kloser, 2009, p. 895)

Complementing the work in physics, in an earlier classroom-based study, Brown and Ryoo (2008) experimentally tested the differential impacts of teaching fifth grade students to understand a scientific construct using everyday language as compared to using only formal scientific language. In the treatment group, concepts and principles of photosynthesis were introduced using everyday language before the introduction of formal scientific language. In the comparison condition, photosynthesis concepts were taught using only formal terminology. Performance on pre-post assessments showed that students in the treatment group developed a deeper understanding of the concepts and principles of photosynthesis (Brown & Ryoo, 2008).

The work of Brown and colleagues (Brown & Kloser, 2009; Brown & Ryoo, 2008) illustrates the complexities and possibilities of recruiting prior knowledge from every-

day experiences and language repertoires from non-academic settings as resources for disciplinary learning. Brown (2019) further developed this argument in his book *Science in the City: Culturally Relevant STEM Education*, wherein he examined how issues of identity, relevance, and perceptions of self-efficacy are taken up in science, technology, engineering, and mathematics (STEM) education that seeks cultural relevance. His arguments are well aligned with the foundational concepts informing the science of human learning and development discussed in this chapter.

Brown's work has important theoretical and empirical implications for instruction and assessment because it demonstrates that attention to conceptual continuities between formal disciplinary discourse and diverse, everyday language has the potential to open up broader learning opportunities than is typically the case. From a theoretical perspective, the results call attention to the need to consider conceptual continuity as a framework for understanding students' science learning. A conceptual continuity framework has the potential to restructure contemporary theories of science learning by challenging the paradigm of teaching from an assessment of "right" and "wrong" answers towards demonstrations of levels of conceptual continuity. This approach reflects a dialogic between concepts as manifest in familiar activities and everyday settings and as manifest in more formal disciplinary contexts.

Attention to relationships among everyday practices and disciplinary knowledge provides opportunities to address the transfer of knowledge directly. Everyday practices, when routine, include a range of kinds of knowledge, including conceptual, procedural, epistemological, and discursive. All four of these dimensions of knowledge are also central to learning in academic disciplines. Making connections between the everyday and the disciplinary and among the multiple dimensions entailed in deep understanding can be embedded into routine practices in classrooms.

### **Case 3: Literacy: Problems of Figuration**

Case 3 illustrates how everyday language and experiences provide students with access to conventions and norms of literary reasoning and interpretation. Specifically, this case examines how to build conceptual and procedural reasoning around problems of figuration in literature. Problems of figuration are uses of language that are not intended to be interpreted literally. Unreliable narration and irony are figurative tropes that may be localized or exist as genres when they characterize the attention of an entire literary work. Both are taken up in everyday discourses and genres (e.g., cartoons, film, music lyrics, visual arts). The reasoning processes entailed in detecting figurative tropes as non-literal and inferred meanings are documented in the world of literary criticism. For example, Wayne C. Booth has written extensively on both approaches in *A Rhetoric of Irony* and *The Rhetoric of Fiction* (Booth, 1975, 1983). Building on literary criticism and the work of George Hillocks (2016), Michael W. Smith developed strategies for teaching students to detect and interpret irony and unreliable narration (Smith, 1989, 1991).

For detecting unreliable narration, Smith extrapolated the following questions as heuristics:

1. Does the narrator's self-interest make you suspicious of his or her reliability?
2. Is the narrator sufficiently experienced to be reliable?

3. Is the narrator sufficiently knowledgeable to be reliable?
4. Is the narrator sufficiently moral to be reliable?
5. Is the narrator too emotional to be reliable?
6. Are the narrator's actions sufficiently inconsistent with his or her words to make you suspicious of his or her reliability?

Through multiple everyday practices, students typically have experience in detecting unreliable narration and extrapolating meaning from such narratives. Smith has conducted several studies in high school classrooms where students are given everyday texts that embody unreliable narration (Smith, 1992). Students then apply Smith's heuristics to these everyday texts as a scaffold to formal literary texts. The *Calvin and Hobbes* cartoon shown in Figure 3-3 is an example of an everyday text used to detect unreliable narration.

Most students will recognize that Calvin, the little boy in the cartoon, really likes the new girl, despite his emotional disputations. This text is accessible, likely of interest to students, and while their reasoning for detecting that Calvin does not mean what he says is likely tacit, it is possible, through dialogue focused on supporting students in making their thinking visible, to help them make explicit the metacognitive reasoning that underlies their recognition of what Calvin really thinks. Such metacognitive reasoning is susceptible to transfer.



FIGURE 3-3 Calvin and Hobbes.

SOURCE: © 1985 Bill Watterson. Reprinted with permission of Andrews McMeel Syndication. All rights reserved.



This teaching approach requires that teachers or curriculum designers are first able to articulate the breadth of knowledge required to tackle the problem of interest—in this case, unreliable narration. They can then identify tasks that recruit students' everyday knowledge, bringing meaningful relations to the disciplinary tasks like understanding sources of congruence and dissonance as well as interactional patterns of discourse, activities, and assessments that provide windows into the developmental pathways from the everyday to the formal discipline.

This case of figuration in literary works draws on students' prior knowledge and cultural repertoires and makes problem solving explicit, thereby supporting students' self-efficacy and sense of relevance. Conceptualizing reasoning processes in everyday practices and their relationships to disciplinary problem solving can inform diagnostic assessments that center relevant knowledge that decontextualized assessments do not.

#### **Case 4: History: Designing for Historical Reading and Reasoning**

Case 4 illustrates a set of instructional design principles that capitalizes on the continuities and discontinuities between everyday language and experiences and disciplinary language and practices. Ms. H, a middle school history teacher, participated in a design-based research project aimed at implementing history instruction that engaged students in developmentally appropriate forms of historical inquiry (Goldman & Popp, 2022; Goldman et al., 2016). The instructional design dealt with several challenges posed by historical reading and reasoning for sixth grade students, including the linguistic complexity of historical documents, students' limited background knowledge of many of the topics and events in the curriculum, and preconceptions about history typically held by students of this age range (Goldman et al., 2016; Lee & Sprately, 2009; National Research Council, 2005). The design principles reflect the developmental principle of balancing what is known with what is new, as well as the importance of making visible what it means to read, think, and reason like a historian. Four instructional strategies were consistently employed throughout the instructional units. Taken together, they built on students' everyday knowledge and forms of linguistic expression to make visible what it means to read, think, and reason like a historian. The four instructional strategies were to:

- *Build on learners' everyday experiences and language* (Lee, 2007; Moll et al., 2006). Historical reasoning practices were first introduced *informally*, using language and experiences that were familiar to students (e.g., Who wrote the article?). More formal labels for historical reasoning practices (e.g., sourcing, corroboration) were introduced only after students were already doing the practice (e.g., taking note of the author, comparing and contrasting content).
- *Make historical reading and reasoning processes visible*. This involved the teacher modeling historical reading and reasoning (i.e., conducting a think-aloud while reading) followed by metacognitive conversations about the modeling. Going "meta" made the teacher's thinking an explicit object of student reflection, thereby increasing their awareness of what the teacher was doing as well as how and why she was doing it. Making these processes visible provided students with

concrete examples of strategies for reading historical texts and ways of thinking that define historical inquiry.

- *Keep complexity manageable by minimizing reading demands when introducing new practices.* For example, when Ms. H first introduced students to the kinds of questions historians ask about artifacts, she did so in the context of objects and photographs. Only after the students had practiced asking these kinds of questions about the objects and photographs were print-based artifacts (e.g., newspaper excerpts, catalog ads) introduced. The same practice was then applied to increasingly more complex and varied text genres.
- *Employ social support for reading linguistically challenging documents and other historical artifacts.* Reading assignments were organized in a sequence of three phases: students independently read and annotated chunks of texts, then discussed with a partner, and then discussed with the whole class (Schoenbach et al., 2012).

These instructional principles were incorporated into a year-long sequence of instructional units that prepared students to conduct their own historical investigations. Ms. H relied on classroom whole and small group discussions, exit slips, and short essays to assess students' thinking throughout the units. These informal assessments showed Ms. H how students were engaging with the historical sources as they debated the merits of claims within those sources. Ms. H attended to what students were noticing in the properties of the sources—in particular, whether they were noting source properties that had implications for interpreting the information contained within (e.g., author, when the source was written, and type of source). Ms. H used this information formatively to make decisions about subsequent lessons. She also regularly modified what she had planned to do the next day to focus on areas where her informal assessments indicated students needed additional opportunities to engage in historical reading and/or reasoning practices. Importantly, Ms. H was attuned to the students' reasoning processes, whether they were expressed through everyday language (e.g., the authors of these two sources are saying different things) or more formal language (e.g., these sources do not corroborate each other's accounts). Over the course of the year, summative assessments encompassed more of the historical inquiry process. That is, early in the year, inquiry tasks for summative purposes might only require that students summarize the position expressed in two different sources, whether the student agreed with that position, and why. Mid-year, summative inquiry tasks required students to evaluate more sources and additional perspectives. By the end of the year, students were provided with more open-ended inquiry tasks and resources from which they could choose what information they would use to provide their descriptive accounts of the focal historical event. Throughout the year, Ms. H downplayed the importance of specific formal terminology (e.g., sourcing, contextualizing, corroborating, chronology, and periodicity) and emphasized the processes to which the terminology refers.

### **Case 5: Recruiting Everyday Repertoires to Support Disciplinary Conceptual, Procedural, and Epistemological Knowledge in Tandem with Identity Development and Engagement: Cultural Modeling**

Cultural Modeling is a design framework aimed at recruiting everyday repertoires to support learning in disciplinary content areas (Lee, 1995, 2007, 2014). Since discourse is essential for learning, engagement, and relating with others, the problem of discourse norms for communication within the classroom is important. In Cultural Modeling, classroom discourse seeks to recruit how students use language and interact with one another to maximize engagement, while simultaneously apprenticing students to understanding and using the language of the discipline orally and in writing. The Cultural Modeling framework draws from syntheses of research on human learning and development that articulate the complex ways that thinking, perception, emotional salience attributed to experience, and relationships work together in learning and development. The framework requires deep analyses of the demands of disciplinary learning; the cultural—including linguistic—repertoires of the discipline and of the learners; and the opportunities that disciplinary learning can offer for identity development. Since neither commercial curriculum nor available assessments capture these multiple dimensions of learning and development, implementing Cultural Modeling in classrooms has historically involved engaging teachers and researchers to collaboratively examine the demands of texts and the prior knowledge and cultural repertoires of their students. These studies were conducted in classrooms and schools that serve predominantly African American student populations that live in low-income communities (e.g., Lee, 1995, 2007). The framework, as developed by Lee (2007) and discussed here, focuses on teaching literary reasoning.

Early work in Cultural Modeling focused on points of convergence between problems of figuration in a genre of African American English called signifying—a form of ritual insult—and in literature (Lee, 1995). Figuration, whether in everyday discourse or works of literature, involves language whose meaning is not literal (e.g., metaphors, symbolism, irony, satire). Everyday knowledge of signifying, as established in sociolinguistics research, entails both reasoning strategies and epistemological dispositions to value figuration. Instructional planning begins by drawing from work in literary criticism to identify established expert heuristics for detecting and interpreting the use of figuration, including symbolism, irony, satire, and unreliable narration. Smith's work (1989, 1991) presented in Case 3 illustrated heuristics for unreliable narration. Once heuristics are identified, planning seeks to identify everyday genres and tasks with which students are familiar and thus are likely to have the skills to interpret. These genres and tasks are referred to as cultural data sets. The first phase of instruction involves students interpreting cultural data sets and engaging in “metacognitive conversations” with their peers during which they make explicit the thinking and reasoning processes they are using. Teachers observe these conversations and support and assist students in explicating their reasoning processes—how they know what they know. Instruction then moves to carefully sequenced literary texts that pose the same problem of figuration with the expectation that students will transfer the processes made visible with the cultural data sets to the literary texts. The Cultural Modeling framework is concerned with both developing technical competence and using disciplinary knowledge as a

medium for “identity wrestling”—an important dimension of human learning and development. Literature offers important opportunities for readers to wrestle with the conundrums of the human experience. Literature focusing on particular cultural communities (race/ethnicity, gender, age cohort, religion, communities at different points in cultural/historical history) entails authors wrestling with life course complexities. In classes that employ Cultural Modeling, the goal is to identify literary texts that offer possibilities for students to grapple with life challenges that are particularly relevant to them as adolescents and members of particular communities.

Cultural Modeling work has been largely carried out with middle- and high-school students—age groups that include important transitional points in adolescent development. This work has also been carried out predominantly with African American students, who must wrestle with both the normative challenges of early and late adolescence and the challenges of navigating and resisting negative stereotypes and structures of discrimination. Thus, the initial formal literary texts in units of instruction invite students to wrestle with issues related to their racial and ethnic identities. Later texts examine similar themes but in different cultural and historical contexts. The classroom design requires that students wrestle with the same technical aspects of figuration first in everyday cultural data sets, then culturally close literary texts, and then culturally distant literary texts.

In Cultural Modeling classrooms, discourse norms recruit how students use language and interact with one another to maximize engagement. Simultaneously, these norms apprentice students into understanding and using the language of the discipline—both orally and in writing. In Cultural Modeling classrooms, when African American students are present, African American English is recruited as a medium of oral communication.

Assessment aligned with the aims of Cultural Modeling addresses the following:

- Everyday knowledge relevant to problem solving in the domain
- Conceptual knowledge in the discipline
- Epistemological knowledge related to the discipline
- Students’ perceptions of learning

Early phases of the Cultural Modeling work included assessments of students’ abilities to interpret signifying dialogues. Students were given assessments of signifying dialogue drawn from exemplars in the sociolinguistic literature (Lee, 1993) as well as assessments of formal literary reasoning based on Hillocks’ taxonomy for assessing literary reasoning (Hillocks & Ludlow, 1984). Hillocks’s taxonomy is an example of how to disentangle processes of comprehension specific to literature. This taxonomy stands in contrast to typical assessments of literature, which pose questions that are outcomes of comprehension but do not provide any windows into the kinds of challenges students face in comprehending and interpreting literature. Hillocks’s taxonomy includes the following:

1. Basic stated information—explicit and central to the narrative.
2. Key details—occur at important points in the narrative and bear causal relationships with what happens in the narrative.

3. Stated relationship—relationship between at least two pieces of information in the narrative.
4. Simple implied relationship—similar to stated relationships except they must be inferred and the details are typically localized within a section of the narrative.
5. Complex implied relationship—relationships that must be inferred; details informing the inference are distributed across the text.
6. Author generalization—questions about themes.
7. Structural generalization—questions about the language and structural choices made by the author and what they imply.

Essentially, Hillocks’s taxonomy provides criteria for different levels of literal and inferential comprehension, as well as broader extrapolation and attention to features of the entire text. As such, it offers a framework for both designing literature comprehension questions and for differentiating among different levels of literary text comprehension. For example, the final two question types in the list above—author generalization and structural generalization—are crucial for literary interpretation. When assessments are designed by teachers, Hillocks’s taxonomy can serve as an instructional planning tool because teachers must analyze for themselves the sources of complexity in literary texts. This kind of qualitative analysis goes beyond traditional measures of text complexity (Goldman & Lee, 2014).

In a three-year longitudinal study in a high school serving African American students living in a low-income community, Lee (2016) included measures of reading based on Hillocks’s taxonomy; epistemological knowledge assessed through a measure of epistemological dispositions toward reading literature (Yukhymenko-Lescroart et al., 2016); self-efficacy; established measures of racial identity (Sellers et al., 1998); and students’ perceptions of learning using the TRIPOD survey (Kuhfeld, 2017), an established and widely used instrument that captures students’ perceptions of learning along seven dimensions:

- Care—show concern for students’ emotional and academic well-being
- Confer—encourage and value students’ ideas and views
- Captivate—spark and maintain student interest in learning
- Clarify—help students understand content and resolve confusion
- Consolidate—integrate and synthesis of key ideas
- Challenge—insist that students persevere and do their best work
- Classroom management—foster orderly, respectful, and on-task classroom behavior

TRIPOD served as a formative assessment, in that it was given as a pre- and post-test each year, providing teachers and the school community with data regarding students’ perceptions of their experiences in English Language Arts classrooms. TRIPOD focused teachers’ attention on the salience of students’ perceptions, and in terms of school climate, revealed opportunities for the department and school administration to consider how to address these important dimensions of learning and engagement.

Examining relations across these multiple measures, researchers found positive relationships between students’ everyday knowledge, conceptual knowledge in liter-



ary reasoning, a positive racial identity, epistemological beliefs in the social functions of reading literature and importance of multiple readings, and positive perceptions of the learning environment and instructional practices (Lee, 2016). The use of multiple measures that index a range of constructs interacting to support learning illustrates a holistic systematic opportunity to understand robust learning.

### **Case 6: Building Teacher Professional Learning Communities as Central to Building Capacity for Learning, Teaching, and Assessments: Chèche Konnen**

Cases 1 through 5 reveal some of the complexities of connecting students' everyday repertoires to disciplinary learning. For teachers to learn to navigate such complexities, they need systemic supports. The knowledge required for such instructional planning and assessment development is complex and not typically embedded in teacher professional development in the United States. There is, of course, the example of the Lesson Study in Japan—where teachers in school-based communities research their own practices—but due to systemic difference, the Lesson Study model has not been tractable in the United States given the organization of teachers' workload and school day (Fernandez & Yoshida, 2004; Lewis et al., 2006).

Case 6 illustrates a model of support for teacher learning through participation in professional learning communities, where teachers and other collaborators examine a problem of practice together. Such collaborative wrestling often yields new insights that arise through dialogic interactions within these communities of practice—insights that are rare when teachers work through these problems of practice individually. Case 6 focuses on a collaboration between the Chèche Konnen professional learning community of teachers, who work with diverse student populations in the Boston area, and researchers who collaboratively investigate the teaching of science and mathematics in diverse classrooms (Warren et al., 2001). In Haitian Creole, Chèche Konnen means, roughly, “to find out.” In this work, they have documented many instructional exemplars that recruit everyday repertoires to support STEM learning. The teams collaborate in planning instruction, but equally important teachers bring to the group problems of practice, and situations where students do and say things that are challenging to fully understand in the moment. Unpacking these situations as a group, with time to reflect, can provide new insights into student thinking and understanding—a key component of assessment situated close to instruction. The work of the Chèche Konnen professional learning community supports teachers as they learn to make in situ evaluations of student activity and discussions.

In Case 6, we focus on one example of a discussion among the Chèche Konnen community about one teacher's unit on plant growth. The teacher's class is ethnically and racially diverse and many students are multilingual, with different degrees of competence in speaking English. The discussion under focus here regards two Latinx students in Mrs. Pertuz's third grade classroom (Ballanger, 2004). One is middle class with parents who are university professionals. The second is a recent immigrant who is dominant in Spanish and emergent in English. The unit being taught focused on understanding the conditions of plant growth. The middle-class and English-dominant Latinx student conveyed the logic of plant growth by referencing a formal chart students created. The teacher understood this student's argument because it mapped onto

the formal representations she had taught in the class. By contrast, the teacher struggled in the moment to understand the explanation provided by Elena, the recent immigrant who was less fluent in English.

Elena: "I think I got the answer to Juana's question. That I don't-I don't think we could see them grow but I think they could feel themselves grow. Sometimes we can feel ourselves grow because my feet grow so fast cuz this little crinkly thing is always bothering my feet. That means it's starting to grow. It's starting to stretch out." (Ballanger, 2004, p. 308)

Teachers cannot fully predict what students will say or do during instruction, and when students' language and/or actions do not map directly to teachers' expectations, they are faced with a conundrum of practice. In this case, Mrs. Pertuz brought this discussion to her Chèche Konnen learning community and together they struggled to understand the logic and epistemological assumptions informing Elena's response (Warrant & Rosebery, 2008; Warren et al., 2001). They looked to the history of science for possible explanations. This type of effort—to continuously think critically about the discipline being taught—is a core requirement for linking everyday prior knowledge and dispositions with those of the academic disciplines in ways that inform instruction and assessment. The group concluded that what has come to be called embodied cognition is and has been a heuristic used by scientists when investigating a phenomenon about which they have limited formal understanding. They focused in on a particular exemplar of embodied cognition as manifest in Albert Einstein's imagining and reasoning about time that were inspired by the clock tower in downtown Bern, Switzerland.

Einstein heard the toll one evening in May 1905. He had been confounded by a scientific paradox for a decade, and when he gazed up at the tower he suddenly imagined an unimaginable scene. What, he wondered, would happen if a streetcar raced away from the tower at the speed of light?

If he was sitting in the streetcar, he realized, his watch would still be ticking. But looking back at the tower, the clock – and time – would seem to have stopped. It was a break-through moment. Six weeks later, he finished a paper outlining a "special theory of relativity." Later he would show how space-time, as he called it, affected mass, energy, and gravity, foreshadowing the nuclear age, space travel, and our understanding of how stars and celestial bodies interact. (Bleiberg, 2016)

Einstein imagined himself inside the phenomenon of interest as a resource for making sense of a phenomenon he did not fully understand. Such positioning has been documented in the history of science as a mode of reasoning when confronting unknown phenomena. The Chèche Konnen professional learning community drew on this embodied reasoning to make sense of Elena's response. Elena was drawing on her own lived experience of knowing that she was growing but not actually being able to see that growth. She imagined that it might be the same for plants. Thus, the group connected Elena's reasoning to Einstein's experience and connected that to findings in the study of embodied cognition.

Mrs. Pertuz, through her dialogic collaborative problem solving in a professional learning community, recognized that Elena was introducing a new epistemological

resource for reasoning about scientific phenomena, particularly where one's formal knowledge may be less clear or developed: the role of imagination—placing oneself inside the phenomenon of interest, just as Einstein had. Mrs. Pertuz was now able to think about how to both scaffold Elena's learning and how to make a wider range of reasoning resources available to her students. In collaboration with her professional learning community, Mrs. Pertuz was able to engage with her class in this broader context instead of simply interpreting Elena's response as incorrect.

Chèche Konnen is one example of professional learning communities that exist in the United States—within schools, within practitioner organizations, and across multiple sites (e.g., The National Writing Project). Chapter 5 of this volume, "Assessment Literacy and Professional Learning," discusses the importance of professional learning communities for teacher learning, as well as the important features and supports needed for such communities, while Chapter 6 of this volume, "District and School Practices and Assessments to Support a Learning-Centered Vision," expands on necessary school and district supports.

Case 6 represents an example of the kind of embedded assessments that teachers routinely conduct during daily instruction. As a field, educators need to understand the range of pedagogical content knowledge, child and adolescent development as relevant in K–12 classrooms, the funds of knowledge that students bring to the classroom from their lived experiences (Moll & Gonzales, 2004), and the importance of such knowledge as a critical element of a balanced assessment system.

### **Case 7: District Level: Research Practice Partnership**

To have the greatest impact at scale, assessment systems need to address relations among teaching and assessments within classrooms and within and across schools. Penuel and Watkins (2019) report on a research-practice partnership (Coburn & Penuel, 2016) involving Denver Public Schools, University of Colorado Boulder, Northwestern University, the Tidemark Institute, Clark University, the Biological Sciences Curriculum Study (BSCS), and Project VOYCE. Key features of this partnership include collaboration at all levels of the educational system and across the project in identifying goals, practices, and evaluation, with equity as an overarching theme. The partnership defined

an equitable educational *system* as one in which all students encounter opportunities where they can connect what they are learning to their lives outside of school and that help them to imagine and pursue futures where they can apply knowledge and practices at work, in civic and family life, and at play. (Penuel & Watkins, 2019, p. 205; Penuel et al., 2016)

In addition, the partnership focused on what they call "epistemic justice," which involves attention to, appreciation of, and uptake of modes of reasoning and foundational belief systems from across diverse communities as central to processes of learning and knowing (Fricker, 2007). This project focused specifically on teaching, how students learn, and assessment in science education.

Important takeaways from this district-level research–practice partnership include:

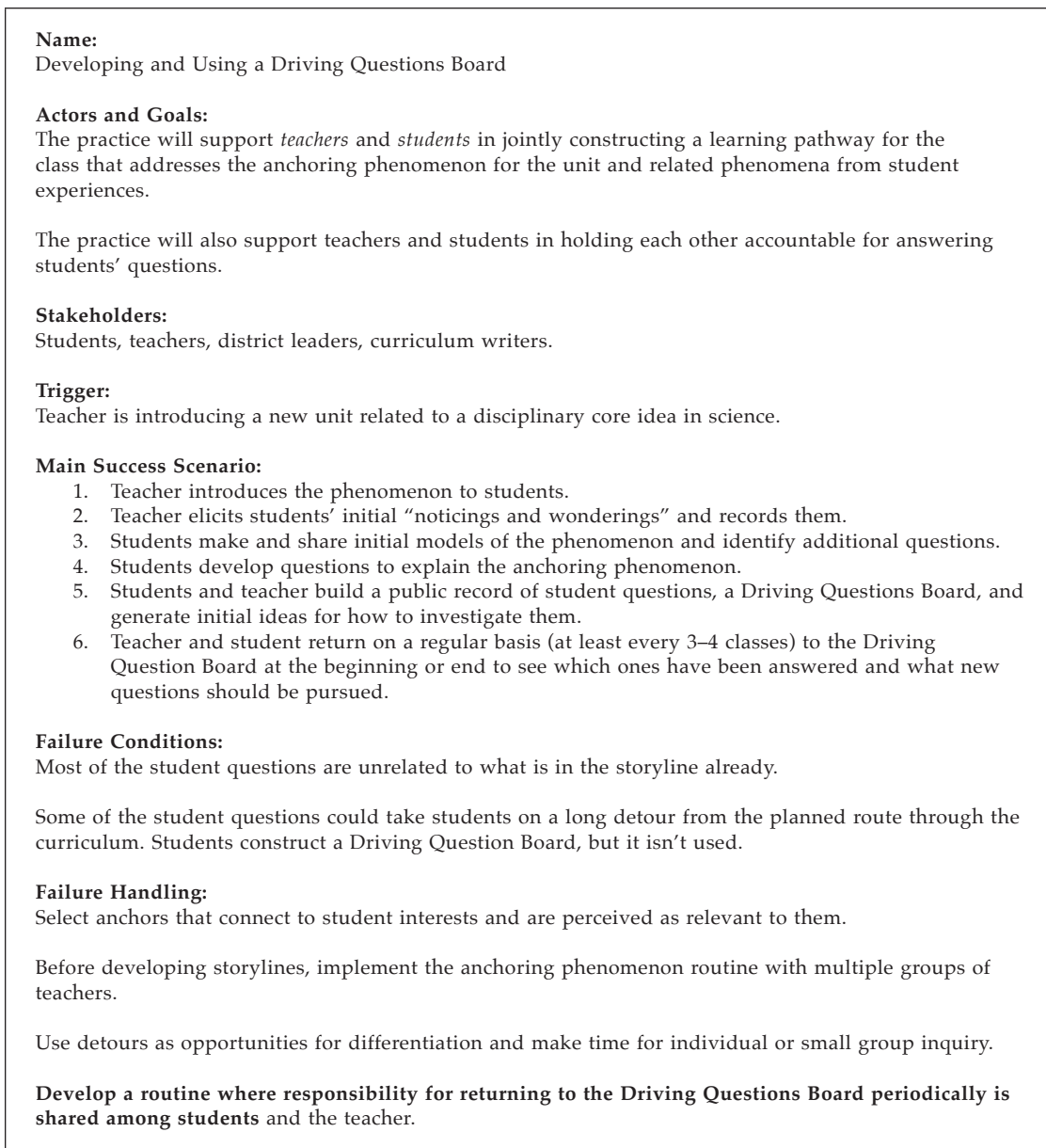
- Core instructional practices were designed through collaborations between district-level leadership, school-level leadership, teachers, and community partners.
- The assessments developed and used focused not only on cognitive outcomes but also importantly on indicators of students' engagement, perceptions of relevance, and efficacy.
- The underlying design of instruction and assessment required collaborative teams to identify investigations in instructional materials that were rooted in students' interests and expectations, as well as to have students identify and lead investigative projects addressing the application of scientific reasoning to a real-world problem.

Two important features of the instructional and assessment design were protocols for developing investigation questions and how teachers would evaluate students' perceptions of their experiences during the course of the investigations. To develop questions for investigations, the partnership created protocols for "anchoring phenomenon routines" to be enacted by teachers. To exemplify the protocols, Figure 3-4 provides the protocol for *Developing and Using a Driving Questions Board* (Penuel & Watkins, 2019).

These routines were collaboratively developed and regularly reviewed as teaching unfolded, and were thus subject to in-process revisions from members of the partnership. The attention in this protocol to identifying failures in implementation and design along with guidelines for how to address them reflects how this practice can serve as an assessment tool.

Evaluating the unfolding investigations included not only teachers and district personnel but also and importantly, students themselves. The project design included the use of Student Electronic Exit Tickets (SEETs) at pivotal points in the unfolding of a unit. The student entries are digital, allowing access and analyses by the collaborative planning groups. Importantly for achieving the goals of equity and epistemic justice, these exit tickets expand on typical exit tickets that ask students to demonstrate purely cognitive understanding of a lesson. SEETs address constructs that reflect many of the characteristics we listed in Box 3-3, such as relevance to students' lives and their communities, students' perceptions of lesson coherence, and students' sense of belonging in science class. (See for discussion Penuel et al., 2023.) Penuel and Watkins (2019) describe the use of these SEETs as follows:

We examine variation in equity of experience and epistemic justice both within classrooms and across classrooms, looking for patterns that show evidence of epistemic injustice (e.g., fewer African American students are contributing to large group discussions or feeling that their voices are consequential in such discussions) as well as to inequity of opportunity (e.g., some teachers are not using the driving question board at all, while others are using it to partner with students in setting the direction for the units). Then, in a meeting that includes district leaders and partners who help us to design and provide professional learning opportunities for teachers, we discuss results and their implications for supporting teachers in ways that can better meet our partnership's goals for equity and epistemic justice. (p. 210)



**FIGURE 3-4** Use Case 1: Building and making use of a driving question board.  
SOURCE: Adapted from Penuel & Watkins (2019). Reprinted with permission.

Case 7 is an exemplar of how the processes of recruiting students' funds of knowledge, supporting students' identities as learners, and the challenges of such work can be supported at scale with deliberative collaborations among key stakeholders—including students themselves—seeking to attend directly to dialogic relations among instruction and assessments. Chapter 6 of this volume, "District and School Practices and Assessments to Support a Learning-Centered Vision," offers further guidance on such collaborations.



### **Case 8: Systemic Work in Assessment: PISA**

The final case illustrates what is involved in creating national systems of teaching and assessment that provide the breadth and depth of data necessary to help understand and explain variation in learning outcomes. Case 8 examines the Programme for International Student Assessment (PISA), an international assessment of 15-year-olds in reading, science, and mathematics. We offer PISA as a contrast to the only national K–12 assessment in the U.S. education sector—NAEP. NAEP assesses reading, mathematics, science, history, and civics in grades 4, 8, and 12 and reports levels of proficiency for knowledge outcomes in these content areas. NAEP also gives surveys to teachers, administrators, and students—in part to capture data on opportunity to learn (e.g., resource allocations, instructional practices) and asks students about their perceptions of each content area. However, the breadth and depth of issues addressed in NAEP surveys are not as expansive as those used in PISA. For example, PISA asks students about their sense of well-being and connections to school. This kind of attention to social and affective well-being reflects dimensions of learning and development discussed in this chapter— dimensions that go beyond attention only to cognitive outcomes.

We note that recent efforts by panels established by the National Assessment Governing Board (NAGB) to spearhead revisions to the next iterations of NAEP assessments in mathematics and reading have called for changes that can have greater explanatory power, including changes to post-test surveys that capture both opportunity to learn and psychosocial variables (e.g., self-efficacy, motivation, and engagement) that correlate with national and sub-group performances. While these recommendations have been accepted by NAGB, how they will be implemented is yet to be seen. We think, therefore, that it is informative to consider how PISA has addressed the assessment of dimensions beyond the cognitive.

Neither NAEP nor PISA focus on individual scores, but rather group trends over time nationally and, in the case of PISA cross-nationally, as a function of periodic administration to targeted population groups. Thus, they not only document performance at varying grade and age levels but also how those performances change over time and their relationships to postsecondary outcomes like participation in higher education and the workforce. They draw from multiple assessments and surveys to extract inferences about longitudinal patterns. However, these inferences are not about the same populations, but due to their size comparisons across data at different time points in the same participating nations, the assessments offer the possibility of inferring broad longitudinal trends.

In addition to assessment results, the PISA 2018 report includes a longitudinal examination of data from the Trends in International Mathematics and Science Study for data on fourth grade students as well as the Survey of Adult Skills, a product of the Organisation for Economic Co-operation and Development (OECD) Programme for the International Assessment of Adult Competencies (Organisation for Economic Co-operation and Development, 2018). The main PISA assessment program for 15-year-olds also includes indicators of students' sense of self-efficacy, sense of belonging in schools, effort and perseverance, career expectations, and measures of both concentrations of economic disadvantage and disciplinary climate in schools (Organisation for Economic Co-operation and Development, 2018). Analyses explore how equity in students' well-

being has evolved as well as the extent to which disadvantaged students are socially and emotionally resilient.

The OECD PISA international assessment of reading, science, and mathematics is given to 15-year-olds in participating nations every four years. Along with reports of proficiency outcomes, OECD also produces a social disparity report. As expressed in the 2018 report, the PISA Social Disparities report examines the complexities of how socioeconomic status impacts learning outcomes across participating nations (Organisation for Economic Co-operation and Development, 2018):

[T]he fact that the impact of social background on educational success varies greatly across countries shows there is nothing inevitable about disadvantaged students performing worse than more advantaged students. Results from education systems as different as Estonia, Hong Kong, Shanghai and Viet Nam show that the poorest students in one region might score higher than the wealthiest students in another country. Within countries too, there are many students who succeed despite predicted failure. On average across OECD countries, more than one in ten disadvantaged students are among the top quarter of achievers in science. (p. 3)

The report concludes:

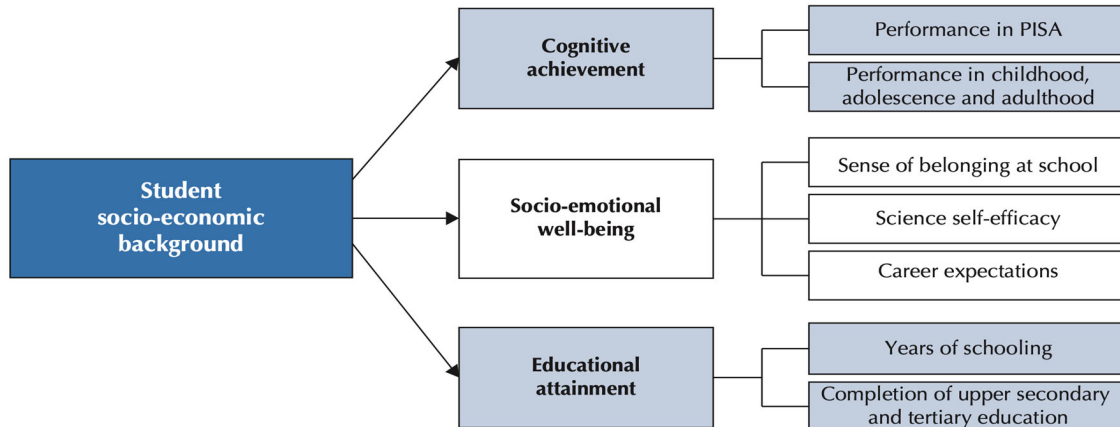
Countries can also set ambitious goals for and monitor the progress of disadvantaged students, target additional resources towards disadvantaged students and schools, and reduce the concentration of disadvantaged students in particular schools. They can also develop teachers' capacity to identify students' needs and manage diverse classrooms, promote better communication between parents and teachers, and encourage parents to be more involved in their child's education. Teachers and schools can foster students' well-being and create a positive learning environment for all students by emphasizing the importance of persistence, investing effort and using appropriate learning strategies, and by encouraging students to support each other, such as through peer-mentoring programmes. (Organisation for Economic Co-operation and Development, 2018, p. 15)

Broadly speaking, OECD takes a broad ecological framing for documenting and understanding trends in social disparities around educational equity, as illustrated in Figure 3-5.

In addition to the indicators outlined in Figure 3-5, OECD draws from extant research to help inform interpretations of findings.

The 2012 PISA report on social disparities includes data and recommendations for policies and practices that build capacity in the educational system to support all students, especially students from low socioeconomic backgrounds (Organisation for Economic Co-operation and Development, 2014). These include supports for teachers, equitable allocation of resources across all schools, and robust pedagogy and curriculum. More details on such systemic supports are discussed in Linda Darling-Hammond's *The Flat World and Education: How America's Commitment to Equity Will Determine Our Future* (2010).

In referencing PISA, we must also acknowledge critiques of the program (Sjøberg, 2016; Teltemann & Klieme, 2017). PISA has been criticized as privileging develop-



**FIGURE 3-5** Equity in education outcomes.  
 SOURCE: Organisation for Economic Co-operation and Development (2018).

ing countries and not adequately addressing issues of cultural relevance of content. However, even with these limitations, OECD’s efforts to address systemic features of educational systems that contribute to PISA outcomes are worth investigating.

We offer this final case as an existence proof that it is possible to design a program of assessment that can both inform needed changes and identify what works in a system—as opposed to assessment programs that only consider cognitive outcomes. This is particularly important because whether from NAEP data or international comparisons from PISA and Progress in International Reading Literacy Study (PIRLS), socioeconomic status and race/ethnicity continue to be associated with disparities in performance outcomes.

### Cross-Case Analysis

We offer these eight cases to illustrate possibilities of how teaching and assessment practices can address the features of robust equitable teaching, learning, and assessments as articulated in Box 3-3. In Table 3-5, we summarize the features of such practices that are captured in each case. These cases provide but a glimpse into the kinds of considerations that need to be taken into account in the design of balanced assessment systems. We would be the first to admit that while each of these cases depicts some important features of assessment *for* or *as* learning, none of them constitute an exemplar of a balanced assessment system.

The cases presented in this chapter are useful in that they are not merely theoretical but have been enacted in real classrooms and schools. At the same time, there are few exemplars of assessment systems that address the goals of equity in opportunity to learn and that seamlessly connect all levels of the system—broader policies, school culture, and classroom climate—and that include instruction and assessment across all levels. We hope these cases stimulate insights and innovative ideas for conceptualizing balance assessment systems that are equitable and just.

**TABLE 3-5** Features of Equitable Teaching, Learning, and Assessment for Each Case

Case	Case	Features of Equitable Teaching and Assessment
1	Mathematics: Examining Everyday Repertoires of Practice as Linked to Disciplinary Learning	<ol style="list-style-type: none"> <li>1. Addresses cognitive processes and practices of reasoning.</li> <li>2. Examines relations between multiple sources of prior knowledge and targets of new learning.</li> <li>3. Makes visible how students reason through mathematical inquiry.</li> </ol>
2	Science: Relationships Among Discourse Registers	<ol style="list-style-type: none"> <li>1. Addresses cognitive processes and practices of reasoning.</li> <li>2. Examines relations between multiple sources of prior knowledge and targets of new learning.</li> <li>3. Makes visible how students reason through science inquiry.</li> </ol>
3	Literacy: Problems of Figuration	<ol style="list-style-type: none"> <li>1. Addresses cognitive processes and practices of reasoning.</li> <li>2. Makes visible how students reason through literary inquiry.</li> </ol>
4	History: Designing for Historical Reading and Reasoning	<ol style="list-style-type: none"> <li>1. Addresses cognitive processes and practices of reasoning.</li> <li>2. Examines relations between multiple sources of prior knowledge and targets of new learning.</li> <li>3. Makes visible how students reason through historical inquiry.</li> </ol>
5	Recruiting Everyday Repertoires to Support Disciplinary Conceptual, Procedural, and Epistemological Knowledge in Tandem with Identity Development and Engagement: Cultural Modeling	<ol style="list-style-type: none"> <li>1. Addresses cognitive processes of reasoning.</li> <li>2. Addresses perceptions of:               <ol style="list-style-type: none"> <li>a. self-efficacy</li> <li>b. mindset</li> <li>c. motivation</li> <li>d. relevance</li> </ol> </li> <li>3. Examines relations between multiple sources of prior knowledge and targets of new learning.</li> <li>4. Makes visible how students reason through literary inquiry.</li> <li>5. Includes learners, teachers, classroom climate, and school culture.</li> </ol>
6	Building Teacher Professional Learning Communities as Central to Building Capacity for Learning, Teaching, and Assessments: Chèche Konnen	<ol style="list-style-type: none"> <li>1. Addresses cognitive processes of reasoning.</li> <li>2. Examines relations between multiple sources of prior knowledge and targets of new learning.</li> <li>3. Makes visible how students reason through problem solving.</li> <li>4. Includes learners, teachers, classroom climate, and school culture.</li> </ol>
7	District Level: Research Practice Partnership	<ol style="list-style-type: none"> <li>1. Addresses cognitive processes and practices of reasoning.</li> <li>2. Examines relations between multiple sources of prior knowledge and targets of new learning.</li> <li>3. Provides actionable data on opportunity to learn versus accountability that negatively impacts students, teachers, and schools.</li> <li>4. Addresses perceptions of:               <ol style="list-style-type: none"> <li>a. self-efficacy</li> <li>b. mindset</li> <li>c. motivation</li> <li>d. relevance</li> </ol> </li> <li>5. Includes learners, teachers, classroom climate, and school culture.</li> </ol>
8	Systemic Work in Assessment: PISA	<ol style="list-style-type: none"> <li>1. Provides actionable data on opportunity to learn versus accountability that negatively impacts students, teachers, and schools.</li> <li>2. Addresses perceptions of:               <ol style="list-style-type: none"> <li>a. self-efficacy</li> <li>b. mindset</li> <li>c. motivation</li> <li>d. relevance</li> </ol> </li> <li>3. Includes learners, teachers, classroom climate, school culture, and district policies.</li> </ol>

## CONCLUSION

There are multiple complex challenges to enabling the vision of balanced assessment systems that are “intentionally designed to provide feedback to students and information for teachers to support ambitious and equitable instructional and learning opportunities” (Chapter 1 of this volume, “Reimagining Balanced Assessment Systems: An Introduction,” p. 2). The challenges are not merely technical. They require a fundamental reconceptualization of human learning and development. They require understanding the multiple pathways through which humans as individuals and communities engage in sense-making, problem solving, and learning. In addition, these challenges call for attention to the multi-dimensional, interactive dialogic processes that contribute to human learning and development. Lastly, these challenges necessitate fundamental reconceptualization of knowledge in the academic domains—not only in terms of cultural practices across diverse communities, but also in the history of how, when, and under what circumstances knowledge in these disciplines evolved and continues to evolve. These reconceptualizations have strong implications for foundational concepts in assessment theory—in particular, validity and how assessment validity is determined. We must ask “Valid for whom, under what circumstances, and in what contexts?” Addressing these challenges will require building infrastructures for professional learning communities among educators—teachers; school administrators; and district, state, and federal leaders—because the commercial resources typically available to schools are restrictive. Robust teaching leading to equitable outcomes cannot be based on curricula that impose scripted teaching and uniform pacing of instructional content. Rather equitable teaching and assessment requires that teachers be adaptive experts (Hatano & Inagaki, 1986) so that they may implement rigorous and challenging instruction that respects and values their students and communities at the same time that it opens up multiple pathways to disciplinary learning.

The processes through which research informs policy also present challenges, including political processes that the research community often does not thoroughly understand. The uptake of the recommendations made in this report is complicated by the fact that public education in the United States is constitutionally the purview of individual states. The current heated battles over what is taught in schools at district and state levels are complex—how to teach history and what is included as part of that discipline, and the banning of books and topics—virtually all of which are influenced by perceptions and belief systems around race/ethnicity and gender/sexual orientation (Pollock et al., 2022). In short, the uptake of the recommendations from this report is not simply a technical exercise.

The field of assessment can offer substantive levers to support robust learning to the extent that assessments can:

- shed light on the multiple dimensions of knowledge, including how these dimensions differ across academic disciplines;
- tap into the psychosocial dimensions of learning (e.g., perceptions of the self, self-efficacy, relevance);
- be sufficiently dynamic to capture multiple pathways and modes of reasoning; and
- address opportunity to learn.



Furthermore, assessment systems should not be limited to formal schooling. Learning takes place in multiple settings, particularly in communities outside of school. Assessment systems should be broad enough to include supports for learning in the variety of non-school settings in which people interact and learn.

In conclusion, for assessment systems to achieve equity, they must be sufficiently flexible to be responsive to the diversity of pathways and funds of knowledge that students from across diverse communities bring to the learning process. This flexibility means that support will be required for all levels of the assessment system, including the work of teachers in classrooms; administrators in schools and districts; and policy makers at district-, state-, and federal levels. Achieving equity also means expanding expectations for learning outcomes beyond limited technocratic goals and that these expectations must address the holistic needs of youth and communities. We have argued that a broad conception of human learning and development—including the cognitive, social, cultural, and identity dimensions that contribute to learning—are captured in current syntheses of the science of human learning and development, a science that takes up propositions from sociocultural theories of learning, but substantially expands understanding of the intertwined nature of these dimensions.

## REFERENCES

- 21st Century Workforce Commission. (2000). *A nation of opportunity: Building America's 21st century workforce*. National Alliance of Business. <https://files.eric.ed.gov/fulltext/ED442971.pdf>.
- Ananiadou, K., & Claro, M. (2009). *21st century skills and competences for new millennium learners in OECD countries*. OECD Education Working Papers. <http://doi.org/10.1787/218525261154>.
- Ascher, M. (1991). *Ethnomathematics: A multicultural view of mathematical ideas*. CRC Press: Taylor & Francis Group.
- Au, K. H. (2013). A sociocultural model of reading instruction: The Kamehameha Elementary Education Program. In S. A. Stahl & D. A. Hayes (Eds.), *Instructional models in reading* (pp. 181–202). Routledge: Taylor & Francis Group.
- Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). Jossey-Bass.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28(2), 117–148. [https://doi.org/10.1207/s15326985ep2802\\_3](https://doi.org/10.1207/s15326985ep2802_3).
- Bang, M., & Marin, A. (2015). Nature–culture constructs in science learning: Human/non-human agency and intentionality. *Journal of Research in Science Teaching*, 52(4), 530–544. <https://doi.org/10.1002/tea.21204>.
- Barron, B. (2006). Interest and self-sustained learning as catalysts of development: A learning ecology perspective. *Human Development*, 49, 193–224. <http://dx.doi.org/10.1159/000094368>.
- Bazerman, C. (1985). Physicists reading physics: Schema-laden purposes and purpose-laden schema. *Written Communication*, 2, 3–23.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy, and Practices*, 18(1), 5–25.
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). Springer.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability*, 21, 5–31.
- Black, R. W. (2005). Access and affiliation: The literacy and composition practices of English-language learners in an online fanfiction community. *Journal of Adolescent & Adult Literacy*, 49(2), 118–128. <https://doi.org/10.1598/JAAL.49.2.4>.

- Booth, W. C. (1975). *A rhetoric of irony*. The University of Chicago Press.
- Booth, W. C. (1983). *The rhetoric of fiction* (2nd ed.). The University of Chicago Press.
- Booy, G., Hendriks, R. J. J., Smulders, M. J. M., Van Groenendael, J. M., & Vosman, B. (2000). Genetic diversity and the survival of populations. *Plant Biology*, 2(4), 379–395. <https://doi.org/10.1055/s-2000-5958>.
- Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (5th ed., Vol. 1, pp. 993–1028). Wiley.
- Brown, B. A. (2019). *Science in the city: Culturally relevant STEM education*. Harvard Education Press.
- Brown, B. A., & Kloser, M. (2009). Conceptual continuity and the science of baseball: Using informal science literacy to promote students' science learning. *Cultural Studies of Science Education*, 4, 875–897. <https://doi.org/10.1007/s11422-009-9198-1>.
- Brown, B. A., & Ryoo, K. (2008). Teaching science as a language: A “content-first” approach to science teaching. *Journal of Research in Science Teaching*, 45(5), 529–553. <https://doi.org/10.1002/tea.20255>.
- Cantor, P., Osher, D., Berg, J., Steyer, L., & Rose, T. (2018). Malleability, plasticity, and individuality: How children learn and develop in context. *Applied Developmental Science*, 1–31.
- Champion, T. (1997). “Tell me somethin’ good”: A description of narrative structures among African American children. *Linguistics and Education*, 9(3), 251–286. [https://doi.org/10.1016/S0898-5898\(97\)90002-4](https://doi.org/10.1016/S0898-5898(97)90002-4).
- Champion, T. (2003). *Understanding storytelling among African American children: A journey from Africa to America*. Erlbaum.
- Coburn, C. E., & Penuel, W. R. (2016). Research-practice partnerships in education: Outcomes, dynamics, and open questions. *Educational Researcher*, 45(1), 48–54. <https://doi.org/10.3102/0013189X16631750>.
- Cole, M. (1998). *Cultural psychology: A once and future discipline*. Harvard University Press.
- Cole, M. (2007). Phylogeny and cultural history in ontogeny. *Journal of Physiology*, 101(4–6), 236–246. <https://doi.org/10.1016/j.jphysparis.2007.11.007>.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Routledge: Taylor & Francis Group.
- Curwood, J. S., Magnifico, A. M., & Lammers, J. C. (2013). Writing in the wild: Writers' motivation in fan-based affinity spaces. *Journal of Adolescent & Adult Literacy*, 56(8), 677–685. <https://doi.org/10.1002/JAAL.192>.
- Damasio, A. R. (1995). REVIEW: Toward a neurobiology of emotion and feeling: Operational concepts and hypotheses. *The Neuroscientist*, 1(1), 19–25. <https://doi.org/10.1177/107385849500100104>.
- Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*. Teachers College Press.
- Darling-Hammond, L., & Conley, D. T. (2015). Assessment systems for deeper learning. In J. Bellanca (Ed.), *Deeper learning: Beyond 21st century skills* (pp. 235–271). Solution Tree Press.
- de Brey, C., Musu, L., McFarland, J., Wilkinson-Flicker, S., Diliberti, M., Zhang, A., Branstetter, C., & Wang, X. (2019). *Status and trends in the education of racial and ethnic groups 2018 (NCES 2019-038)*. National Center for Education Statistics, U.S. Department of Education.
- Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching*, 46(6), 606–609. <https://doi.org/10.1002/tea.20316>.
- Duschl, R., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: A review and analysis. *Studies in Science Education*, 47(2), 123–182. <https://doi.org/10.1080/03057267.2011.604476>.
- Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109–132.
- Elby, A., Richards, J., Walkoe, J., Gupta, A., Russ, R. S., Luna, M. J., Robertson, A., Coffey, J. E., Edwards, A. R., Sherin, M., & van Es, E. A. (2014). Differing notions of responsive teaching across mathematics and science: Does the discipline matter? In J. L. Polman, E. A. Kyza, D. Kevin O’Neill, I. Tabak, W. R. Penuel, A. Susan Jurow, K. O’Connor, T. Lee, & L. D’Amico (Eds.), *Learning and becoming in practice: The International Conference of the Learning Sciences (ICLS) 2014: Volume 2* (pp. 1406–1415). International Society of the Learning Sciences. [https://archive.isls.org/conferences/icls/2014/downloads/ICLS%202014%20Volume%202%20\(PDF\)-wCover.pdf](https://archive.isls.org/conferences/icls/2014/downloads/ICLS%202014%20Volume%202%20(PDF)-wCover.pdf).

- Emdin, C. (2013). The rap cypher, the battle, and reality pedagogy: Developing communication and argumentation in urban science education. In M. L. Hill & E. Petchauer (Eds.), *Schooling hip-hop: Expanding hip-hop based education across the curriculum* (pp. 11–27). Teachers College Press.
- Evans, R. W. (Ed.). (2021). *Handbook on teaching social issues* (2nd ed.). Information Age Publishing.
- Fernandez, C., & Yoshida, M. (2004). *Lesson study: A Japanese approach to improving mathematics teaching and learning*. Routledge: Taylor & Francis Group.
- Feuerstein, R. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques*. University Park Press.
- Feuerstein, R., Feuerstein, R. S., Falik, L. H., & Rand, Y. (2002). *The dynamic assessment of cognitive modifiability: The learning propensity assessment device: Theory, instruments and techniques*. International Center for the Enhancement of Learning Potential.
- Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
- Fullan, M. (2015). Breakthrough learning. In J. Bellanca (Ed.), *Deeper learning: Beyond 21st century skills*. Leading Edge.
- Gee, J. P. (1992). *The social mind: Language, ideology, and social practice*. Bergin and Garvey.
- Ginsburg, H. P., & Opper, S. (1988). *Piaget's theory of intellectual development* (3rd ed.). Prentice-Hall, Inc.
- Goldman, S. R., & Lee, C. D. (2014). Text complexity: State of the art and the conundrums it raises. *The Elementary School Journal*, 115(2), 290–300. <https://doi.org/10.1086/678298>.
- Goldman, S. R., & Popp, J. S. (2022). Transforming sixth-grade social studies from “just the facts” to historical inquiry: A case study of teacher learning. *The History Teacher*, 56(1), 44–75.
- Goldman, S. R., & Snow, C. (2015). Adolescent literacy: Development and instruction. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 463–478). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199324576.013.2>.
- Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., Lee, C. D., Shanahan, C., & Project READI. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist*, 51(2), 1–28. <https://doi.org/10.1080/00461520.2016.1168741>.
- Good, C., & Dweck, C. S. (2006). Motivational orientations that lead students to show deeper levels of reasoning, greater responsibility for their academic work, and greater resilience in the face of academic difficulty. In R. J. Sternberg & R. F. Subotnik (Eds.), *Optimizing student success in school with the other three Rs: Reasoning, resilience, and responsibility* (pp. 39–56). Information Age Publishing.
- Gopnik, A., Meltzoff, A. N., & Kuhl, P. (1999). *The scientist in the crib: Minds, brains, and how children learn*. William Morrow & Co.
- Greeno, J. G., & Middle School Mathematics through Applications Project Group (1998). The situativity of knowing, learning, and research. *American Psychologist*, 53(1), 5–26. <https://doi.org/10.1037/0003-066X.53.1.5>.
- Griffin, P., McGaw, B., & Care, E. (2012). *Assessment and teaching of 21st century skills*. Springer.
- Gutiérrez, K. D., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, 32(5), 19–25.
- Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. W. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262–272). W. H. Freeman.
- Heider, F. (1958). *The psychology of interpersonal relations*. John Wiley & Sons Inc. <https://doi.org/10.1037/10628-000>.
- Hillocks, G. (2016). The territory of literature. *English Education*, 48(22) 109–126.
- Hillocks, G., & Ludlow, L. H. (1984). A taxonomy of skills in reading and interpreting fiction. *American Educational Research Journal*, 21(1), 7–24. <https://doi.org/10.3102/00028312021001007>.
- Holland, D., & Lave, J. (2009). Social practice theory and the historical production of persons. *Actio: An International Journal of Human Activity Theory*, 2, 1–15.
- Immordino-Yang, M. H., & Damasio, A. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, Brain, and Education*, 1(1), 3–10. <https://doi.org/10.1111/j.1751-228X.2007.00004.x>.
- Jaber, L. Z., Dini, V., & Hammer, D. (2022). “Well that’s how the kids feel!”—Epistemic empathy as a driver of responsive teaching. *Journal of Research in Science Teaching*, 59(2), 223–251. <https://doi.org/10.1002/tea.21726>.

- Knorr Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge*. Harvard University Press.
- Kuhfeld, M. (2017). When students grade their teachers: A validity analysis of the Tripod student survey. *Educational Assessment*, 22(4), 253–274. <https://doi.org/10.1080/10627197.2017.1381555>.
- Kuhl, P. K. (2011). Early language learning and literacy: Neuroscience implications for education. *Mind, Brain, and Education*, 5(3), 128–142. <https://doi.org/10.1111/j.1751-228X.2011.01121.x>.
- Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, 100(4), 2425–2438.
- Kuhl, P. K., Ramírez, R. R., Bosseler, A., Lin, J.-F. L., & Imada, T. (2014). Infants' brain responses to speech suggest analysis by synthesis. *Proceedings of the National Academy of Sciences*, 111(31), 11238–11245. <https://doi.org/10.1073/pnas.1410963111>.
- Kuhn, T. (2012). *The structure of scientific revolutions*. The University of Chicago Press.
- Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal*, 27(1), 29–64. <https://doi.org/10.3102/00028312027001029>.
- Lampert, M. (2003). *Teaching problems and the problems of teaching*. Yale University Press.
- Lave, J. (2012). Changing practice. *Mind, Culture, and Activity*, 19(2), 156–171. <https://doi.org/10.1080/10749039.2012.666317>.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Harvard University Press.
- Lee, C. D. (1993). *Signifying as a scaffold for literary interpretation: The pedagogical implications of an African American discourse genre*. National Council of Teachers of English. <https://files.eric.ed.gov/fulltext/ED357371.pdf>.
- Lee, C. D. (1995). A culturally based cognitive apprenticeship: Teaching African American high school students skills in literary interpretation. *Reading Research Quarterly*, 30, 608–631.
- Lee, C. D. (2007). *Culture, literacy, and learning: Taking bloom in the midst of the whirlwind*. Teachers College Press.
- Lee, C. D. (2010). Soaring above the clouds, delving the ocean's depths: Understanding the ecologies of human learning and the challenge for education science. *Educational Researcher*, 39(9), 643–655. <https://doi.org/10.3102/0013189X10392139>.
- Lee, C. D. (2014). A multi-dimensional cultural modeling framework for the design of robust literacy instruction. In P. J. Dunston, S. K. Fullerton, M. W. Cole, D. Herro, J. A. Malloy, P. M. Wilder, & K. N. Headley (Eds.), *63rd yearbook of the Literacy Research Association* (pp. 78–85). Literacy Research Association.
- Lee, C. D. (2016). *A longitudinal study of disciplinary literacy development in literature and history as a resource for identity development and psycho-social well-being*. Project READI. [https://projectreadi.org/wp-content/uploads/2017/04/READI-Tech-Report-11\\_LongitudinalStudyDisciplinaryLiteracyDev.pdf](https://projectreadi.org/wp-content/uploads/2017/04/READI-Tech-Report-11_LongitudinalStudyDisciplinaryLiteracyDev.pdf).
- Lee, C. D. (2017). Integrating research on how people learn and learning across settings as a window of opportunity to address inequality in educational processes and outcomes. *Review of Research in Education*, 41(1), 88–111. <https://doi.org/10.3102/0091732X16689046>.
- Lee, C. D., & Sprately, A. (2009). *Reading in the disciplines: The challenges of adolescent literacy*. Carnegie Corporation of New York. <https://www.carnegie.org/publications/reading-in-the-disciplines-the-challenges-of-adolescent-literacy>.
- Lee, C. D., Meltzoff, A. N., & Kuhl, P. K. (2020). The braid of human learning and development: Neurophysiological processes and participation in cultural practices. In N. S. Nasir, C. D. Lee, R. Pea, & M. McKinney de Royston (Eds.), *Handbook of the cultural foundations of learning* (pp. 24–43). Routledge: Taylor & Francis Group.
- Lee, C. D., White, G., & Dong, D. (Eds.). (2021). *Educating for civic reasoning and discourse*. National Academy of Education.
- Legette, K. (2018). School tracking and youth self-perceptions: Implications for academic and racial identity. *Child Development*, 89, 1311–1327. <https://doi.org/10.1111/cdev.12748>.
- Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? A case of lesson study. *Educational Researcher*, 35(3), 3–14. <https://doi.org/10.3102/0013189X035003003>.
- Lucas, S. B. (1999). *Tracking inequality, stratification, and mobility in American high schools*. Teachers College Press.



- Marion, S. (2020). *Using opportunity-to-learn data to support educational equity*. National Center for the Improvement of Educational Assessment. [https://www.ncea.org/wp-content/uploads/2021/11/CFA-Marion.OTL\\_Indicators.pdf](https://www.ncea.org/wp-content/uploads/2021/11/CFA-Marion.OTL_Indicators.pdf).
- Martin, A. J., & Marsh, H. W. (2006). Academic resilience and its psychological and educational correlates: A construct validity approach. *Psychology in the Schools, 43*(3), 267–281. <https://doi.org/10.1002/pits.20149>.
- Meltzoff, A. N. (1988). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology, 24*(4), 470–476.
- Meltzoff, A. N. (2013). Origins of social cognition: Bidirectional self-other mapping and the “Like-Me” hypothesis. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us* (pp. 139–144). Oxford University Press.
- Meltzoff, A. N., & Kuhl, P. K. (2016). Exploring the infant social brain: What’s going on in there? *Zero to Three Journal, 36*(3), 2–9.
- Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science, 325*(5938), 284–288. <https://doi.org/10.1126/science.1175626>.
- Michalchik, V., & Gallagher, L. (2010). Naturalizing assessment. *Curator: The Museum Journal, 53*(2), 209–219. <https://doi.org/10.1111/j.2151-6952.2010.00020.x>.
- Moll, L. C., & Gonzáles, N. (2004). Engaging life: A funds-of-knowledge approach to multicultural education. In J. A. Banks & C. A. McGee Banks (Eds.), *Handbook of research on multicultural education* (2nd ed.). Jossey-Bass.
- Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (2006). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. In N. Gonzalez, L. Moll, & C. Amanti (Eds.), *Funds of knowledge: Theorizing practices in classrooms, communities, and households* (pp. 83–100). Routledge.
- Nasir, N. S., Lee, C. D., Pea, R., & McKinney de Royston, M. (Eds.). (2020). *Handbook of the cultural foundations of learning*. Routledge.
- Nasir, N. S., Lee, C. D., Pea, R., & McKinney de Royston, M. (2021). Rethinking learning: What the interdisciplinary science tells us. *Educational Researcher, 50*(8), 557–565. <https://doi.org/10.3102/0013189X211047251>.
- National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. The National Academies Press. <https://doi.org/10.17226/24783>.
- National Assessment of Educational Progress. (n.d.). *Explore NAEP long-term trends in reading and mathematics*. The Nation’s Report Card. <https://www.nationsreportcard.gov/ltr/?age=9>.
- National Center for Education Statistics. (2009). *NAEP 2008: Trends in academic progress*. U.S. Department of Education. [https://nces.ed.gov/nationsreportcard/pdf/main2008/2009479\\_1.pdf](https://nces.ed.gov/nationsreportcard/pdf/main2008/2009479_1.pdf).
- National Center for Education Statistics. (2013). *A first look: 2013 mathematics and reading: National Assessment of Education Progress at grades 4 and 8*. U.S. Department of Education. <https://nces.ed.gov/nationsreportcard/subject/publications/main2013/pdf/2014451.pdf>.
- National Center for Education Statistics. (2022). *U.S. adults with low literacy and numeracy skills: 2012/14 to 2017*. <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2022004>.
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academy Press. <https://doi.org/10.17226/9853>.
- National Research Council. (2005). *How students learn: History, mathematics, and science in the classroom* (pp. 31–77). The National Academies Press. <https://doi.org/10.17226/10126>.
- National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. The National Academies Press. <https://doi.org/10.17226/13398>.
- Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38*(1/2), 1–135. <https://doi.org/10.2307/1165788>.
- Oakes, J. (2005). *Keeping track: How schools structure inequality*. Yale University Press.
- Organisation for Economic Co-operation and Development. (2014). *PISA 2012 results in focus: What 15-year-olds know and what they can do with what they know*. <https://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf>.
- Organisation for Economic Co-operation and Development. (2018). *Equity in education: Breaking down barriers to social mobility*. OECD Publishing. <https://doi.org/10.1787/9789264073234-en>.



- Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What “ideas-about-science” should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Teaching*, 40(7), 692–720. <https://doi.org/10.1002/tea.10105>.
- Osher, D., Cantor, P., Berg, J., Steyer, L., & Rose, T. (2018). Drivers of human development: How relationships and context shape learning and development. *Applied Developmental Science*, 1–31.
- Penuel, W. R., & Shepard, L. A. (2016). Assessment and teaching. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of Research on Teaching* (5th ed., pp. 787–850). American Educational Research Association.
- Penuel, W. R., & Watkins, D. A. (2019). Assessment to promote equity and epistemic justice: A use-case of a research-practice partnership in science education. *The ANNALS of the American Academy of Political and Social Science*, 683(1), 201–216. <https://doi.org/10.1177/0002716219843249>.
- Penuel, W. R., Clark, T. R., & Bevan, B. (2016). Infrastructures to support equitable STEM learning across settings. *Afterschool Matters*, 24, 13–20.
- Penuel, W. R., Krumm, A. E., Pazera, C., Singleton, C., Allen, A.-R., & Deverel-Rico, C. (2023). Belonging in science classrooms: Investigating their relation to students’ contributions and influence in knowledge building. *Journal of Research in Science Teaching*, 61, 228–252. <https://doi.org/10.1002/tea.21884>.
- Pinkard, N. D. (1999). Lyric reader: An architecture for creating intrinsically motivating and culturally responsive reading environments. *Interactive Learning Environments*, 7(1), 1–30. <https://doi.org/10.1076/ilee.7.1.1.3599>.
- Pollock, M., Rogers, J., Kwako, A., Matschiner, A., Kendall, R., Bingener, C., Reece, E., Kennedy, B., & Howard, J. (2022). *The conflict campaign: Exploring local experiences of the campaign to ban “Critical Race Theory” in public K–12 education in the U.S., 2020–2021*. Institute of Democracy, Education, and Access, University of California, Los Angeles. <https://idea.gseis.ucla.edu/publications/files/the-conflict-campaign-report>.
- Quartz, S. R., & Sejnowski, T. J. (2002). *Liars, lovers, and heroes: What the new brain science reveals about how we become who we are*. William Morrow.
- Rogoff, B. (1997). Evaluating development in the process of participation: Theory, methods, and practice building on each other. *Change and development: Issues of theory, method, and application* (pp. 265–285). Routledge: Taylor & Francis Group.
- Rogoff, B. (2003). *The cultural nature of human development*. Oxford University Press.
- Rogoff, B., & Chavajay, P. (1995). What’s become of research on the cultural basis of cognitive development. *American Psychologist*, 50(10), 859–877. <https://doi.org/10.1037/0003-066X.50.10.859>.
- Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers’ informal formative assessment practices and students’ understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44(1), 57–84. <https://doi.org/10.1002/tea.20163>.
- Schoenbach, R., Greenleaf, C., & Murphy, L. (2012). Reading for understanding: How reading apprenticeship improves disciplinary learning in secondary and college classrooms. Jossey-Bass.
- Sellers, R. M., Smith, M. A., Shelton, J. N., Rowley, S. A., & Chavous, T. M. (1998). Multidimensional model of racial identity: A reconceptualization of African American racial identity. *Personality and Social Psychology Review*, 2(1), 18–39. [https://doi.org/10.1207/s15327957pspr0201\\_2](https://doi.org/10.1207/s15327957pspr0201_2).
- Shanahan, C., Shanahan, T., & Mischia, C. (2011). Analysis of expert readers in three disciplines: History, mathematics, and chemistry. *Journal of Literacy Research*, 43, 393–429.
- Shepard, L. A. (2019). Classroom assessment to support teaching and learning. *Annals of the American Academy of Political and Social Science*, 683, 183–200.
- Shepard, L. A. (2021). Ambitious teaching and equitable assessment: A vision for prioritizing learning, not testing. *American Educator*, Fall. <https://www.aft.org/ae/fall2021/shepard>.
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practices*, 37(1), 21–34. <https://doi.org/10.1111/emip.12189>.
- Sjøberg, S. (2016). OECD, PISA, and globalization: The influence of the international assessment regime. In C. H. Tienken & C. A. Mullen (Eds.), *Education policy perils: Tackling the tough issues*. Routledge: Taylor & Francis Group.
- Smith, M. W. (1989). Teaching the interpretation of irony in poetry. *Research in the Teaching of English*, 23(3), 254–272.

- Smith, M. W. (1991). *Understanding unreliable narrators: Reading between the lines in the literature classroom*. National Council of Teachers of English. <https://files.eric.ed.gov/fulltext/ED331091.pdf>.
- Smith, M. W. (1992). Effects of direct instruction on ninth graders' understanding of unreliable narrators. *The Journal of Educational Research*, 85(6), 339–347.
- Smitherman, G. (1995). "Students' right to their own language": A retrospective. *The English Journal*, 84(1), 21–27. <https://doi.org/10.2307/820470>.
- Spencer, M. B. (2006). Phenomenology and ecological systems theory: Development of diverse groups. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 829–893). John Wiley & Sons, Inc.
- Taylor, E. V. (2009). The purchasing practice of low-income students: The relationship to mathematical development. *Journal of the Learning Sciences*, 18(3), 370–415. <https://doi.org/10.1080/10508400903013462>.
- Teltemann, J., & Klieme, E. (2017). The impact of international testing projects on policy and practice. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 369–386). Routledge: Taylor & Francis Group.
- Tomasello, M. (2021). *Becoming human: A theory of ontogeny*. Harvard University Press.
- Toulmin, S., Rieke, R., & Janik, A. (1984). *An introduction to reasoning*. Macmillan Publishing Company.
- Tyson, K. (2013). Tracking segregation, and the opportunity gap. In P. Carter & K. G. Welner (Eds.), *Closing the opportunity gap: What America must do to give every child an even chance* (pp. 169–180). Oxford University Press.
- Uccelli, P., Galloway, E. P., Barr, C. D., Meneses, A., & Dobbs, C. L. (2015). Beyond vocabulary: Exploring cross-disciplinary academic-language proficiency and its association with reading comprehension. *Reading Research Quarterly*, 50(3), 337–356.
- Van de Veer, R. (2014). *Lev Vygotsky*. Bloomsbury Publishing.
- Voogt, J., & Pareja Roblin, N. (2012). A comparative analysis of international frameworks for 21st century competencies: Implications for national curriculum policies. *Journal of Curriculum Studies*, 44, 299–321.
- Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.
- Warren, B., Ballenger, C., Ogonowski, M., Rosebery, A. S., & Hudicourt-Barnes, J. (2001). Rethinking diversity in learning science: The logic of everyday sense-making. *Journal of Research in Science Teaching*, 38(5), 529–552. <https://doi.org/10.1002/tea.1017>.
- Werner, H. (1957). *The concept of development*. University of Minnesota Press.
- William, D. (2011). *Embedded formative assessment*. Solution Tree Press.
- William and Flora Hewlett Foundation. (2013). *Deeper learning competencies: April 2013*. [https://hewlett.org/wp-content/uploads/2016/08/Deeper\\_Learning\\_Defined\\_April\\_2013.pdf](https://hewlett.org/wp-content/uploads/2016/08/Deeper_Learning_Defined_April_2013.pdf).
- Yukhymenko-Lescroart, M. A., Briner, S. W., Magliano, J. P., Lawless, K., Burkett, C., McCarthy, K. S., Lee, C. D., & Goldman, S. R. (2016). Development and initial validation of the Literature Epistemic Cognition Scale (LECS). *Learning and Individual Differences*, 51, 242–248. <https://doi.org/10.1016/j.lindif.2016.09.014>.
- Zaretskii, V. K. (2009). The zone of proximal development: What Vygotsky did not have time to write. *Journal of Russian & East European Psychology*, 47(6), 70–93. <https://doi.org/10.2753/rpo1061-0405470604>.

# Classroom Activity Systems to Support Ambitious Teaching and Assessment

Maria Araceli Ruiz-Primo, *Stanford University*

Erin Marie Furtak, *University of Colorado Boulder*

## CONTENTS

INTRODUCTION .....	94
CONCEPTUALIZING AMBITIOUS TEACHING .....	94
CONCEPTUALIZING AMBITIOUS CLASSROOM ASSESSMENT .....	95
LEARNING ENVIRONMENTS THAT SUPPORT AMBITIOUS TEACHING AND ASSESSMENT .....	97
CLASSROOM ACTIVITY SYSTEM: AN ORGANIZATIONAL FRAMEWORK FOR CLASSROOM TEACHING AND ASSESSMENT .....	97
Learners.....	99
Curriculum .....	103
Instruction.....	108
Assessment .....	111
Classroom Learning Culture.....	116
BRINGING IT TOGETHER: IMPLICATIONS FOR THE DESIGN OF CLASSROOM ASSESSMENT TO SUPPORT AMBITIOUS INSTRUCTION .....	119
Culturally and Socially Relevant Assessments .....	119
Fairness and Representation .....	120
Cognitive Demands.....	121
CONCLUSION .....	122
REFERENCES .....	123

## INTRODUCTION

Ms. T's first and second graders are partway through a unit investigating plant growth and development by exploring pumpkins (Rosebery et al., n.d.; Warren & Rosebery, 2011). After germinating seeds *out* of soil in Petri dishes using moist paper towels, Ms. T (a White European American teacher) begins to shift students to exploring how roots grow using a root chamber—a glass-sided container that makes root growth visible in soil (Warren & Rosebery, 2011). Simon (an African American second grader) interrupts her and asks the question, “Did you put magic beans in there or something?” (Warren & Rosebery, 2011, p. 100).

Rather than close down Simon's question, Ms. T seizes on his contribution as a powerful intellectual opportunity. She asks Simon to “say more,” an instructional move that has important implications for both Simon's and the class's learning. Reflecting on the Petri dishes, Simon replies that “he was wondering how seeds could germinate without soil” (Warren & Rosebery, 2011, p. 101). His explanation allows Ms. T to see Simon's question from a different perspective, highlighting the contradiction Simon saw between seeds growing in and out of the soil. Ms. T followed up by exploring these ideas with Simon and drawing on the implications of Simon's question with the whole class.

This interaction, like many others in Ms. T's classroom, was not purely serendipitous. Rather, it was shaped through the intersection of classroom activities, a classroom culture in which students feel safe to share their ideas, and teaching moves in which Ms. T creates—and seizes on—opportunities to explore student ideas and draw implications with the whole class (e.g., Rosebery et al., 2016). By inviting Simon to share and develop his thinking, Ms. T “opened a space for him to shape an identity as a powerful, engaged, and critical scientific thinker – in his own eyes, her eyes, and the eyes of his classmates” (Warren & Rosebery, 2011, pp. 101–102).

## CONCEPTUALIZING AMBITIOUS TEACHING

Ms. T's lesson on the pumpkin life cycle illustrates several salient characteristics of what has come to be called *ambitious teaching*. Many instructional approaches have been described under the umbrella of ambitious teaching (Hammond, 2021; Lampert et al., 2013; Shepard, 2021; Smith et al., 2001; Smylie & Wenzel, 2006; Windschitl et al., 2018), including:

1. Centering the interests and experiences of students from diverse cultural and linguistic backgrounds;
2. Engaging students in rich, authentic tasks with scaffolds to support their participation;
3. Inviting students to be active co-constructors of and participants in their learning through productive classroom discourse that involves reasoning, explaining, analyzing, and justifying;
4. Developing students' disciplinary knowledge and practice in a community of learners; and
5. Utilizing assessments designed and enacted in alignment with these goals.

As Ms. T's response to Simon illustrates, questions and other talk moves that invite students to participate in discussions and build on what peers have said (Michaels et al., 2016) support her in learning more about what her students are thinking. Questions and talk moves draw together threads from prior class conversations, readings, student experiences, and class investigations to help her use those ideas to revise and improve a common representation. The instance described at the start of this chapter—an occasion of informal, everyday assessment—was the intersection of a carefully planned sequence of lessons, a curated set of resources to support student learning, and a commitment to developing and refining students' thinking through ongoing classroom discourse. In this type of environment, students develop their understanding of both foundational science concepts and science practices.

Ambitious teaching principles are grounded in sociocultural theories of learning, situate learning in a cultural context organized by tools and routines, and conceptualize learning as changing participation in disciplinary practices (Chapter 3 of this volume, "Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems"; Brown et al., 1989; Engeström, 2001; Greeno, 2006; Lave & Wenger, 1991). A sociocultural view contemplates classrooms as organized disciplinary communities of practice that attend to the interconnected cognitive, social, emotional, and cultural facets of learning and development. A sociocultural view also encourages consideration of both teachers and students as key participants in classrooms, bringing their previous knowledge, identities, and lived experiences into these learning environments (see Chapter 3 of this volume, "Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems"; Gutiérrez & Rogoff, 2003).

Sociocultural perspectives emphasize that what students learn, think, and feel is the result of complex interactions that reflect their cultural and contextual circumstances (see Chapter 3 of this volume, "Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems"). Learning, then, can be defined "as the transformation of an individual's participation in valued social and cultural activities" (see Chapter 3 of this volume, "Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems," p. 57). Any disciplinary practice involves a set of activities (special ways of acting and interacting to produce and use knowledge; Gee, 2008) and experiences (special ways of seeing, valuing, and being in the world; Gee, 2008). Interactions between learners and their environments that involve those disciplinary activities and experiences evolve, reflecting a change in the characteristics of the participation of learners (e.g., the way to *do* science, to *talk* science, to *value* science; Gee, 2008) and diverse levels of appropriation of the disciplinary practices. In the classroom context, a sociocultural perspective should respond to the diversity of students' home and community cultures. It should be a step toward bridging the gap between the classroom and students' homes and communities (Ladson-Billings, 2021).

### CONCEPTUALIZING AMBITIOUS CLASSROOM ASSESSMENT

As the views of teaching shift, so must the ways that classroom assessment is theorized, designed, and enacted (see Chapter 3 of this volume, "Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems"). Thus, ambitious classroom assessment is integrated into and overlaps with ambitious teaching



practices (Shepard, 2021). A sociocultural perspective has implications for how to think about supporting students' learning and how activities related to assessing students do more than just provide information about students: they shape understanding about what learning is, what is important to learn, and who the learners are (Haertel et al., 2008).

From a sociocultural perspective, assessment means observing, documenting, and analyzing how students use and modify their knowledge, skills, and engagement in disciplinary practices over time to participate in a classroom community (Moss, 2008; National Research Council, 2001). It follows that classroom assessments should be designed with learners' interests and identities in mind. In addition to assessing an in-depth understanding of key knowledge and skills within a domain, teachers should assess learners' engagement in disciplinary practices. Therefore, ambitious assessment—like ambitious teaching—should involve the ways of acting on, interacting with, seeing, and valuing the disciplinary world.

Classroom assessment is based on the idea that much of what teachers and students do in their classrooms can be utilized as evidence of students' learning. Assessment, then, is a part of social interactions and is a socially situated activity (Jordan & Putz, 2003). What students say, write, do, and produce are potential sources of evidence of learning and evidence toward achieving rich learning goals.

In the classroom context, learners and teachers are both participants in assessment. Teachers design and/or select assessment tasks and can also take on an unexpected student question as an assessment opportunity. Teachers must make sense of all information sources about students' developing understanding and engagement in practice, and in turn, make decisions based on this information. Assessment also expands beyond individual learners to include their interactions with each other and their reflections on their learning. Participants' engagement in assessment is defined not only by the tasks embedded in curricula but also by the opportunities that arise from the regular participation of the members of a community that supports ambitious teaching and learning. Assessment events are aligned with what students are doing and learning at any given moment during instruction.

The goals for ambitious learning frame gathering or eliciting evidence about students' learning, as well as analyzing and interpreting that information to inform subsequent instructional actions. These assessment activities can happen informally, at any moment during instruction, or more formally, at specific times (Ruiz-Primo & Furtak, 2006, 2007). This means that classroom assessment involves planned as well as unplanned events that should be viewed as opportunities to learn and refine students' conceptual understanding and disciplinary practices.

Classroom assessment can be both formative—conducted on a day-to-day basis while learning is in process—or summative—conducted at the end of an instructional period (e.g., a unit). Engaging learners in assessment—both formative and summative—provides them with opportunities to assess themselves and their peers and to receive or provide feedback. These activities help learners develop and internalize criteria that define what counts as evidence of their learning and also serve as agents of their own learning. In ambitious teaching classrooms, teachers and students work together to promote the learning of the community.

## LEARNING ENVIRONMENTS THAT SUPPORT AMBITIOUS TEACHING AND ASSESSMENT

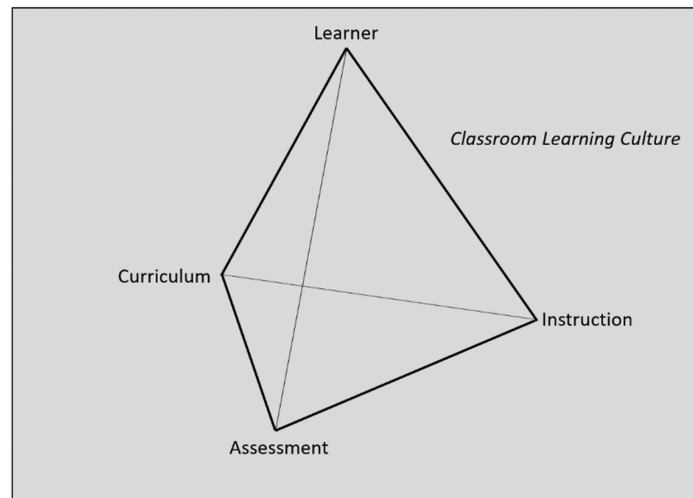
The descriptions of ambitious teaching and ambitious assessment previously presented are part of multifaceted classroom learning environments. As illustrated by the example of Ms. T, these environments reflect a particular culture in which students' learning develops through classroom practices and circumstances (Gay, 2018; Rogoff, 2003). Learners flourish in a classroom culture where everyone contributes; their ideas are valued; and they are supported academically, socially, and emotionally. Learning environments should be informed by the cultures and identities of the learners they serve. In this way, classroom learning environments can be culturally responsive—inviting and building on “the cultural knowledge, prior experiences, frames of reference, and performance styles of ethnically diverse learners to make learning encounters more relevant to and effective for them” (Gay, 2018, p. 36).

Culturally responsive classrooms acknowledge the legitimacy of the cultural heritage of different ethnic groups (attitudes and approaches to learning); build bridges between students' homes and school experiences; use a variety of instructional strategies; teach students how to praise their own and other's cultures; and incorporate multicultural information, materials, and resources into teaching and learning (Gay, 2018). They are conducive to *equitable and collaborative practices* that have the potential to provide all students with multiple opportunities to explain their thinking and develop purposeful and deliberate disciplinary practice. Equitable classroom environments (1) provide access to high-quality instruction and opportunities to learn to all students; (2) offer opportunities to refine students' conceptual understanding and solutions to problems and make connections with students' cultural background and identity; (3) support an expanded view of disciplinary knowledge and practice (e.g., what counts as science); and (4) support teachers and students in seeing disciplinary work as part of justice movements (National Academies of Sciences, Engineering, and Medicine, 2022).

In this chapter, we build on sociocultural theories of learning to define a classroom activity system that supports ambitious teaching and classroom assessment. The activity system is comprised of multiple elements, each informed by what research says about how students learn and the kinds of classroom practices that support their learning. In the next sections, we focus on each of the elements of the activity system. This chapter dives deeply into each element—learners, curriculum, instruction, learning culture, and assessment—and provides examples for each. We start with the critical element, the learners, and then discuss curriculum, instruction, assessment, and the classroom learning culture. Finally, we describe how the elements work together to create a learning environment that supports ambitious classroom assessment to support all learners.

## CLASSROOM ACTIVITY SYSTEM: AN ORGANIZATIONAL FRAMEWORK FOR CLASSROOM TEACHING AND ASSESSMENT

We begin with a framework for a classroom activity system (previously proposed by Ruiz-Primo, 2021), presented in Figure 4-1. The framework illustrates classroom activity systems as an interplay of five elements: *learner, curriculum, instruction, and assessment*, all immersed in a *classroom learning culture*. These five elements have an interdependent purpose in supporting student learning. The framework reflects the



**FIGURE 4-1** Classroom activity system elements.  
SOURCE: Ruiz-Primo, 2021.

relationship among teachers and learners (the who), the subject matter reflected in a curriculum (the what), the instructional and assessment approaches (the how), and the environment in which this all happens (the where).

In this chapter, we describe research-based practices that support ambitious teaching and assessment by attending to learners’ cultural and personal backgrounds, as well as ways of knowing, doing, and being as resources to support their learning (Hammond, 2021; Shepard, 2021; Windschitl et al., 2018). While this chapter’s ultimate purpose is to highlight the role of assessment in a classroom environment that supports ambitious teaching, we also use the activity system to define how assessment is embedded within a larger system of learners, instruction (and teachers), curriculum, and a classroom learning culture.

The framework reflects multiple orientations (Lee, 2008) encompassed by a socio-cultural perspective, including:

1. A *sociocognitive orientation* that reflects the importance of helping learners develop their knowledge and abilities progressively over time (National Research Council, 2000; Penuel & Shepard, 2016; Tobias & Everson, 2009; Zimmerman & Moylan, 2009);
2. A *distributed view of cognition* that reflects the importance of interactions among people and with tools in the classroom, as well as interactions among people with tools and tasks to construct learning (Lee, 2008; Newman et al., 1984); and
3. A *cultural orientation* that acknowledges multiple ways of knowing; considers learners’ background, prior knowledge, and “funds of knowledge” (i.e., accumulated developed bodies of knowledge and skills in households); and the norms and routines that guide interactions in classrooms (Gee, 2008; McDermott & Pea, 2020; Moll et al., 1992; Moss, 2008; Nasir et al., 2020; National Academies of Sciences, Engineering, and Medicine, 2018b; Rosado-May et al., 2020; Ruiz-Primo et al., 2022).

## Learners

Learners come to school with prior knowledge and lived experiences from their homes, families, and communities. This knowledge influences what is seen, heard, and felt, and what learners understand and construct (National Research Council, 2000). In essence, “One’s existing knowledge serves as the foundation of all future learning by guiding organizations and representations, and by coloring and filtering all new experiences” (Murphy & Alexander, 2007, p. 16). Human development is a cultural process and learners’ development largely depends on their social environments, including their norms and relational characteristics (Rogoff, 1995, 2003). Therefore, learners become members of multiple social environments and cultures (e.g., home, school, neighborhood, or race/ethnicity group). These memberships affect what and how they think and learn.

Individual cognition develops through social interactions, both in and out of school (Alexander, 2006; Penuel & Shepard, 2017; Rogoff & Angelillo, 2002; Rogoff et al., 2003). The roots of learning and development depend on sociocultural interactions. What any individual can come to know is determined by the social collective. Conversely, the collective is a result of a group of learners with diverse prior knowledge and experiences (Brown et al., 1989; Rogoff, 1998).

Learners’ prior knowledge is also developed through school experiences, defined by specific curricula and opportunities to learn—including the kinds of resources and instructional practices they were exposed to—and supported by opportunities to demonstrate what they learned and how that learning is applied. These experiences are likely to have influenced their beliefs about their abilities to complete tasks, intrinsic motivation to learn, strategies to process information and control their own learning, and the way they interact with their classmates. There is evidence that learners’ encounters with diverse cultural classroom contexts influence how they perceive and experience themselves (e.g., self-appraisal and self-esteem) as well as their attributional processes about the self (e.g., *I can learn mathematics* versus *my capacity does not allow me to learn mathematics*) (Rosenholtz & Simpson, 1984; Spencer, 1999).

A crucial component of ambitious teaching and learning is centering the interests of all learners, particularly those from historically marginalized backgrounds, in classroom learning environments (National Academies of Sciences, Engineering, and Medicine, 2022). We focus on three aspects of learners below: their interests and identities, linguistic and cultural capital, and knowledge about themselves as learners.

### *Learners’ Interests and Identities*

Ambitious learning environments begin by activating learners’ natural interests and curiosities and using them as entry points to sequences of learning. Learners are motivated, work harder, persist longer, and learn better when what they are learning seems useful and connected to their motivations, identities, and future goals (National Academies of Sciences, Engineering, and Medicine, 2018b). In the classroom, students develop their identity as learners, which is shaped by the culture established in the classroom.

At the same time, learners do not simply learn *about* content; they also learn ways of *being* (Bruner, 1996; Gutiérrez & Rogoff, 2003). For example, learners develop certain ways of participating within their families, their circle of friends in the neighborhood, and their circle of friends in the school. Learners, then, navigate diverse cultural in and out of school practices that require diverse repertoires or ways of participation (Nasir et al., 2022). These experiences lead to particular ways of talking or participating in each context. Finding ways to connect these skills with academic disciplinary practices can positively affect the development of learners' interests, identities, and performance. Here it is important to mention that classrooms cannot support identity "without embracing the differences in the classroom as resources for learning" (Steel, 2012, p. 1,127). Discovering small differences in social relations can make a big difference in the level of learners' engagement in school (Erickson & Mohatt, 1982).

In our research, we have observed teachers' specific strategies to gather information about their learners. On the first days of the school year, an elementary school teacher asked her learners to write her a letter about themselves, describing what they liked or disliked and whatever else they wanted to share (Ruiz-Primo et al., 2022). This information helped the teacher gain insights about—in learners' own words—things she could connect to and leverage during instruction. For example, if some learners liked basketball, she could use that information as context for teaching mathematics concepts (see Nasir, 2007, for an example of how to use basketball in teaching statistics). Once these interests are considered during instruction, they can also be used in a corresponding assessment task, whether informally (e.g., in a classroom conversation), or formally (e.g., in a test at the end of a unit) (Randall et al., 2021).

### *Learners' Linguistic and Cultural Capital*

Learners' home knowledge and languages can be a foundation for classroom instruction and assessment (Brown et al., 1989; Fine & Furtak, 2020; Lee, 2008; Mehan, 2008; Moll et al., 1992; Shepard, 2021). When learners' home languages differ from the dominant culture, these non-dominant language varieties can be devalued and racialized in the classroom (Flores & Rosa, 2015). This devaluation affects learners' participation in classroom discourse and, therefore, their opportunities to learn (Lee, 1995; Mehan, 2008). These differences can be repositioned as an asset, reflecting the "multi-competence" that multilingual youth bring to the classroom, as they have a broad conception of language and cultural knowledge that, with the appropriate curricular and instructional support, enables them to participate, contribute, and succeed (National Academies of Sciences, Engineering, and Medicine, 2018a). Warren and Rosebery (2011) reflected on the interaction between Ms. T and Simon:

Viewed culturally and historically, Simon's ways with words were neither random nor mysterious. He was speaking from within a powerful intellectual and expressive tradition of African American discourse practices, which includes incisive argumentation, metaphorical invention, counterfactual reasoning, and language play (Lee, 2007; Mitchell-Kernan, 1981; Smitherman, 1977, 2000)... In fact, the language use practices of African American students are frequently misread in school as signs of confusion, off-topic digressions, disengagement or disrespect (Foster, 1983; Michaels, 1981). (p. 101)



Ms. T was receptive to Simon's multicompetence, affirming his cultural identity. Teacher–student language patterns that are closer to students' home and cultural interaction patterns are more successful in improving learning than language patterns that are culturally incongruent to the students (Au & Jordan, 1981; Erickson & Mohatt, 1982; Mohatt & Erickson, 1981). If a teacher does not recognize how learners' everyday ways of expressing ideas reflect disciplinary understandings, they "may fail to capitalize on rich, meaningful opportunities for children's learning" (National Academies of Sciences, Engineering, and Medicine, 2022, p. 96). The "interactional etiquette" (Erickson & Mohatt, 1982, p. 135) students bring to the classroom from their everyday experiences varied from one cultural context to another (e.g., the role of "silence"). Understanding these differences should allow teachers to interpret students' remarks and behaviors and make the necessary adaptations in the classroom (like in the case of Simon).

This reframing of multilingual learners' expertise as a resource for assessment is also illustrated by Khisty and Chval (2002), who provide the example of a teacher of Latino students. The teacher introduced the concept of quadrilateral, asking learners to listen carefully to the word and repeating it more than once: "Qua-dri-lat-er-al, Qua-dri-lat-er-al, Qua-dri-lat-er-al" (p. 158). She then asked the learners whether the sound of the word or which part of the word was something that they recognized. One said "cuadro," to which the teacher responded, "What is a cuadro?" (Khisty & Chval, 2002, p. 158). The discussion led to a co-constructed definition: "cuadros" had a square shape and four sides. This teacher capitalized on learners' knowledge of Spanish by connecting the concept of a quadrilateral to cuadro, and other students learned a new word in Spanish. "What is a 'cuadro'?" is a question that can be considered an informal assessment prompt, or task, to find out more about learners' understanding and build from that knowledge.

Learners also bring multiple ways of knowing and being to the classroom. Close collaboration with students, families, and community members can richly inform curriculum and assessment. For example, Indigenous learners view themselves as part of—and not separate from—nature (Bang & Marin, 2015). Instead of limiting what "counts" as ways of knowing to White, Western epistemologies, learning environments should be constructed in ways that not only honor but invite in students' identities and ways of knowing as foundational elements of their learning (Tzou et al., 2019, 2021). The Learning in Places curriculum, for example, begins with activities that invite students into conversation with their families, then encourages taking nature walks to allow for reflections in school as starting points for conversations about the socio-ecological systems near their homes (Learning in Places Collaborative, n.d.).

Students, families, and community members can both contribute to developing curricula and assist in thinking differently about how to develop assessment tasks. Earnest and colleagues (2023) used classroom observation and family interview data to define the types of tasks that would appropriately assess students in an urban public school that actively engaged with and valued the surrounding community. The analysis of classroom observations and interviews led to the identification of themes that could improve assessment practices by focusing on tasks that were developed using the students, families, and community's "funds of knowledge."

### *Learners' Knowledge About Themselves*

It is important to know what students know about themselves as learners, including the strategies they use to study and learn, how they self-regulate their learning, and how they respond to and use feedback. Learners who are reflective, have appropriate self-regulation strategies, and take control of their actions depending on where they are in their understanding (e.g., ask for help when needed, ask for clarifications) will perform better in school and in general throughout their lives (Alexander, 2006; Murphy & Alexander, 2007; Pugh et al., 2000; Weinstein & Mayer, 1986; Winne, 1995). Strategies are learner-initiated actions and strategic learners are “invested learners” (Alexander, 2006; Palmer & Goetz, 1988).

Strategic thinking should be nurtured. Teachers can encourage learners to think about their thinking—metacognition—to better know what they should do next (Alexander, 2006; Winne & Azevedo, 2014). Being metacognitive means that learners self-monitor their knowledge, can self-direct their actions to improve their knowledge, and self-regulate their learning to become independent and strategic learners. Metacognition and motivation support each other—hence the importance of using activities and topics that connect with students’ interests and are familiar to them. Strategy use is heavily influenced by motivation, which in turn may be affected by previous experiences that have led to learners’ holding positive or negative self-perceptions such as, for example, being a poor student (Palmer & Goetz, 1988).

In the classroom, teachers can support strategic thinking and metacognition during instruction and assessment. For example, competent learners have an ample repertoire of general (e.g., how to study) and domain-specific strategies (e.g., rehearsal strategies to learn multiplication tables). Teachers can ask learners to share and reflect upon the strategies they use for certain tasks, and can help learners to reinforce the strategies by making their characteristics explicit (e.g., organizational strategies for complex tasks may imply planning or outlining steps) and making connections with task characteristics (e.g., simple tasks like naming the names of planets may require less sophisticated strategies than more complex tasks like planning and conducting a scientific investigation). Teachers can provide further support to learners by checking for comprehension failures (e.g., self-questioning whether students are understanding) and by developing strategies that can help them in their learning. For example, teachers can ask students to underline or highlight what they consider important ideas and justify their importance, ask themselves questions about the information they are learning, or organize that information in a manner that helps them connect to what was previously learned in the unit or in the course. Talking aloud while solving a problem, analyzing a literary paragraph, or reading the instructions on how to conduct a scientific investigation can all support students’ metacognition (e.g., “Hm, I do not understand this word, I will circle it and make sure I know what it means before I continue” or “I need to check my calculation before I continue”). Teachers can also model positive self-talk (e.g., “I can do this”) and point out when negative self-talk does not help students to move forward (e.g., “This is too difficult, I cannot respond to these questions”). These strategies can be particularly helpful to lower-performing students, whose thinking may be self-deprecating (e.g., “I will never learn this stuff, it is too complicated”).

For example, a mathematics teacher asked her learners whether they wanted to respond to an easy problem, a medium-difficult problem, or a difficult problem about a certain topic during the warm-up segment of the lesson (Ruiz-Primo et al., 2015). She always chose the difficulty that the majority of the learners selected, and then asked the students why they chose that level of difficulty that day. Some days, the responses involved explanations related to their level of understanding—those days sparked a short discussion about their ability to judge their own understanding or difficulties perceived by the students. These conversations promoted self-monitoring. The teacher was careful to ask after learners whose voices were not heard at first. On occasion, she changed the level of the task—usually from medium to difficult—based on the learners’ responses, challenging them to challenge themselves.

Teachers can seek to know their students’ strengths as learners and what their areas of challenge are (Conley, 2018). Important assessment questions can be answered by intentionally observing students’ behavior: Do most of the learners need to hear explanations more than once before they can discuss them? In any given task, how do students manage their time? What learning strategies do they use?

Self-assessment can also help learners reflect on their learning skills (Conley, 2018). For example, they can reflect on the process of completing a particular product or piece of work. Students can respond to questions provided by the teacher like: “Did I manage my time efficiently to finish this task? Did I get stuck on something while conducting this task? If so, what did I do? Did I rethink my approach to the task? Did I ask for help? How did my actions affect the quality of the product?” Questions like these can help learners think about the type of learners they are and support them in taking control of their learning.

Taken together, these three aspects—learners’ interests and identities, linguistic and cultural capital, and knowledge of themselves as learners—can inform approaches to centering learners in classroom environments. Doing so opens space for them to try new ways of knowing and being and builds motivation and engagement (National Academies of Sciences, Engineering, and Medicine, 2018b, 2019).

## Curriculum

Curriculum sets the rigorous, authentic, and challenging tasks that support learning, teaching, and assessment in the classroom. It is a crucial starting point for building bridges to students’ interests and prior experiences to create more equitable learning opportunities. While the curriculum may be just one of many instructional resources present in a classroom activity system, it is a critical one (Remillard & Kim, 2020). Intended or planned curriculum provides the specificity and organizational structure that guides instruction and assessment (Schmidt et al., 2001). The *intended/planned curriculum* helps teachers understand what, when, and how students have opportunities to learn; have clarity about how the different components of the curriculum fit together; see how the sequence of topics and activities build on each other, making the *enacted or implemented curriculum* more likely to succeed; and have clarity about where to focus assessment of students’ learning—that is, the *learned curriculum* (Giamellaro et al., 2017; Remillard & Kim, 2020; Ruiz-Primo, 2016; Schmidt et al., 2001).

Instructional materials reflect curriculum developers' theories of learning. From a sociocultural perspective, curriculum materials should support not just students' acquisition of knowledge but also center and respond to their lived experiences while seeking to sustain their linguistic resources and cultural practices (Paris, 2012). Meaningful learning opportunities develop in part from demanding and challenging learning goals, but also from how the enactment of the curriculum secures active participation that provides a sense of belonging (Shepard, 2021). In this section, we discuss how curriculum can be designed and adapted to center learners' experiences, knowledge, and identities, as well as what, why, and how teachers need to enact curriculum to support learners.

### *Responding to and Sustaining Learners' Knowledge and Practice*

Curricula for ambitious teaching are designed in ways that respond to and seek to sustain learners' knowledge, practices, cultures, and languages (Ladson-Billings, 1995, 2014; Paris, 2012). Ambitious curricula start with learners in mind, dedicating time to finding out students' interests and building on these interests and home knowledge (e.g., Bang & Medin, 2010; Penuel et al., 2019). This type of curriculum engages learners as active participants and promotes motivation and interest. Curricula can be designed to be *adaptive*—lessons can be easily adjusted by the teacher according to the students' interests; *responsive*—lessons provide opportunities for teachers to respond to students' ideas, cultural backgrounds, and experiences; and *sustainable*—engaging students' evolving linguistic and cultural practices (Paris, 2012).

For example, Bang and Medin (2010) engaged Menominee community members and teachers in designing learning experiences in which the Menominee's "ideas, their public expressions, and the practices and behaviors of individuals and groups" (p. 1,014) were an integral part of learning. These learning experiences engage and intertwine students' everyday experiences with their subject matter learning (Tzou et al., 2019). Another example of culturally sustaining curricula is *Math in a Cultural Context: Lessons Learned from Yup'ik Eskimo Elders*, a supplemental mathematics curriculum developed by Lipka et al. (2005) using an expert-apprentice model that is familiar to Yup'ik students. They combined discourse structures with mathematical content based on students' cultural knowledge and spatial abilities. Results from a randomized controlled experiment conducted in Alaska showed that the "Picking Berries" (representing and measuring) and "Going to Egg Island" (grouping and place value) modules significantly improved students' mathematics performance, with relatively robust effect sizes (0.82 and 0.39; Kisker et al., 2012; see Box 4-1).

### *Curriculum Materials Structure and Sequence*

The nature of the materials to be learned matters for both teachers and learners (Choppin et al., 2021; Remillard & Kim, 2020; Ruiz-Primo et al., 2013; Wang et al., 2013). Materials should be rich, challenging, and organized in a way that facilitates learning (e.g., in an appropriate sequence, with characteristics that can help learners recognize patterns, with appropriate cues that tell learners how to connect and use the

**BOX 4-1**  
**An Example of a Unit Based on the Yup'ik Culture**

**“Going to Egg Island: Adventures in Grouping and Place Values”**

A first- and second-grade interdisciplinary unit on Yup'ik culture, geography, and biology was founded on the lands, everyday practices, and cultural and traditional knowledge of the Yup'ik Eskimo. The module is based on the life of a second-grade girl living in a Yup'ik Eskimo community. Using the girl's experiences, students learn to use the Yup'ik abacus and play traditional Yup'ik games while grouping objects in a variety of ways and investigating number patterns until they have a strong sense of grouping and place values.

Students are taught to communicate orally, using traditional Yup'ik ways of counting using the human body (hands, feet, limbs, and the whole body). Students are provided with ways of counting in other cultures (e.g., Native Americans of the Great Plains and Zulu [South Africa]). The unit includes a letter to the students' families that explains what students will be doing and how the family can support the student at home.

**Math in Cultural Contexts: Lessons Learned from Yup'ik Eskimo Elders**

**The Curriculum**

- Follow the Yup'ik way of knowing by using expert-apprentice modeling. Elders and the teacher first demonstrate a concept to the students (the apprentices). Students begin to approximate the knowledge of the expert, which establishes a collaborative classroom setting.
- Engage students *cognitively* by using analytic creative and practical strategies, *socially* by working together, and *practically* by applying or investigating mathematics problems from their daily lives.
- Promote student collaboration in solving challenging problems that can lead to understanding underlying mathematics principles and procedures.
- Allow different learning modalities, assuming that not all students learn in the same way. The curriculum has hands-on activities based on real-world problems, as well as abstracting and deducing activities using analytic, creative, and practical abilities.
- Promote communication among peers in ways that strengthen students' mathematical and logical thinking and help to understand the reasoning and mathematical decisions of their peers. The materials provide strategies to guide students' conversations, improving how students focus on mathematical thinking and help students support their conceptual understanding by practicing in the context of particular problems.
- Sustain the Yup'ik language. The materials include the Yup'ik words used to describe mathematical concepts along with mathematical terms.
- Promote Yup'ik values in each module. Elders counsel against waste and value listening, learning, working hard, being cooperative, and passing knowledge to others.

**The Assessments**

- Are embedded within instructional activities.
- Require teachers to carefully observe, listen to, and challenge their students' thinking.
- Involve students keeping a daily journal that documents their work as well as a record of their increasing mathematical knowledge and ability to communicate what they know. Students can define, explain, sketch, design, ask questions, revisit them, etc. by using this tool.
- Require teachers to adapt instruction based on the information collected through journals, observation, and listening to students in whole classroom dialogues as well as small group conversations.

SOURCE: Lipka (2003).



information, and with linguistic characteristics that are suitable to diverse students; Giamellaro et al., 2017; Ruiz-Primo, 2016; Shepard, 2021; Windschitl et al., 2018).

Curriculum materials that support ambitious learning are designed to provide multiple pathways for students and teachers. For example, OpenSciEd learning resources take a storyline approach, beginning with an anchoring phenomenon that engages learners and encourages them to draw on their prior experiences to ask questions (Edelson et al., 2021). A storyline approach to sequencing curriculum materials prioritizes coherence from a learner’s perspective—that is, rather than building knowledge sequentially, as an expert might conceive, it assembles pieces in ways that logically respond to sequences of learners’ wonderings and questions (Reiser et al., 2021). A curriculum storyline—such as the one shown in Figure 4-2—connects a series of routines that are a planned part of the enacted curriculum, combining multiple rounds of investigations and assembling pieces of what has been learned so far, what remains to be figured out, and culminating with students developing answers to the questions that were posed at the outset of the unit.

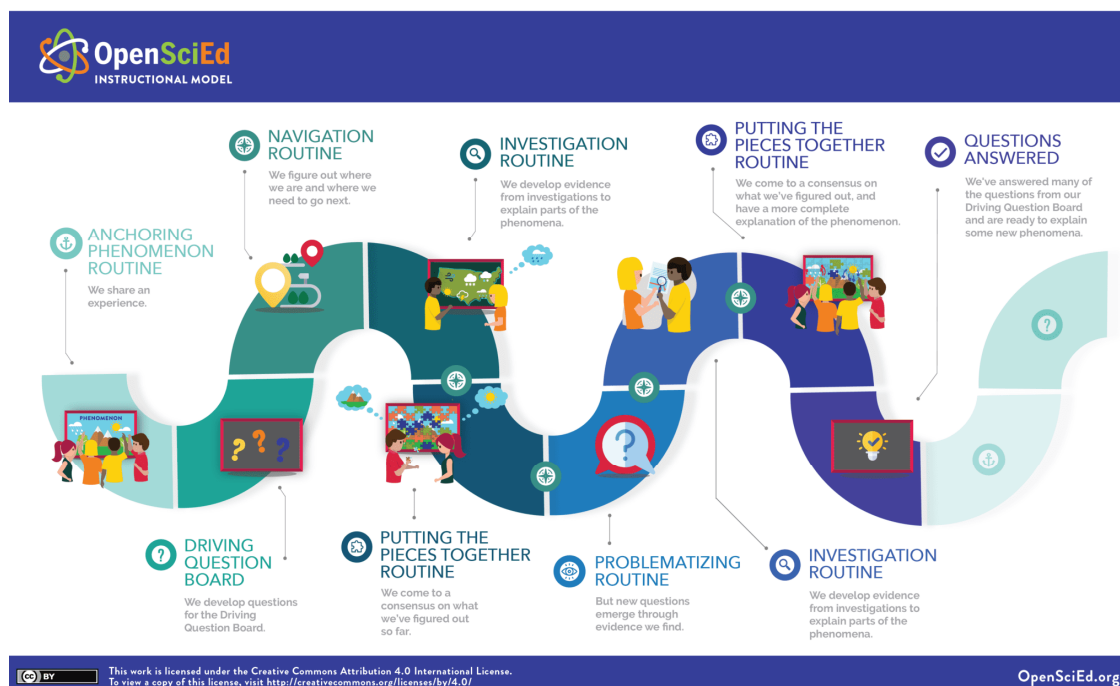


FIGURE 4-2 OpenSciEd storyline model.

SOURCE: OpenSciEd. (n.d.) *Instructional Model*. <https://www.openscienced.org/openscienced-instructional-model>. This work is licensed under a Creative Commons Attribution 4.0 License. <http://creativecommons.org/licenses/by/4.0>.

Curriculum materials that support ambitious teaching and learning are also designed around issues that students experience every day. Figure 4-3 shows the Classroom Storyline developed by Learning in Places, a curriculum that was co-developed with families, educators, and community-based organizations. Learning in Places provides a culturally and community-relevant field-based learning experience for students. Activities within the curriculum explore socio-ecological systems in students' neighborhoods by taking "Wondering Walks," making observations, asking "should we" questions, modeling data and relationships, conducting investigations, analyzing and explaining data, and shared decision making with families and friends. Curricula like Learning in Places allow teachers and students to work on issues that matter to students' everyday lives. What they do impacts not only the students' learning but their families as well (Learning in Places Collaborative, 2023).

Students are motivated to engage in these curricula by questions that engage students' natural curiosity and connect with their lives. For example, a question that guides Learning in Places is, "What do we notice from our Wondering Walks at school and with our families?" Even if curricula are not designed in this manner, teachers can make necessary adaptations so that the curricula are more engaging (deBarger et al., 2017). Curricula such as OpenSciEd and Learning in Places provide students with opportunities to approach problems they encounter in their environment with what they are learning, which makes them more prepared "to be effective members of society" (Ladson-Billings, 2021, p. 7).

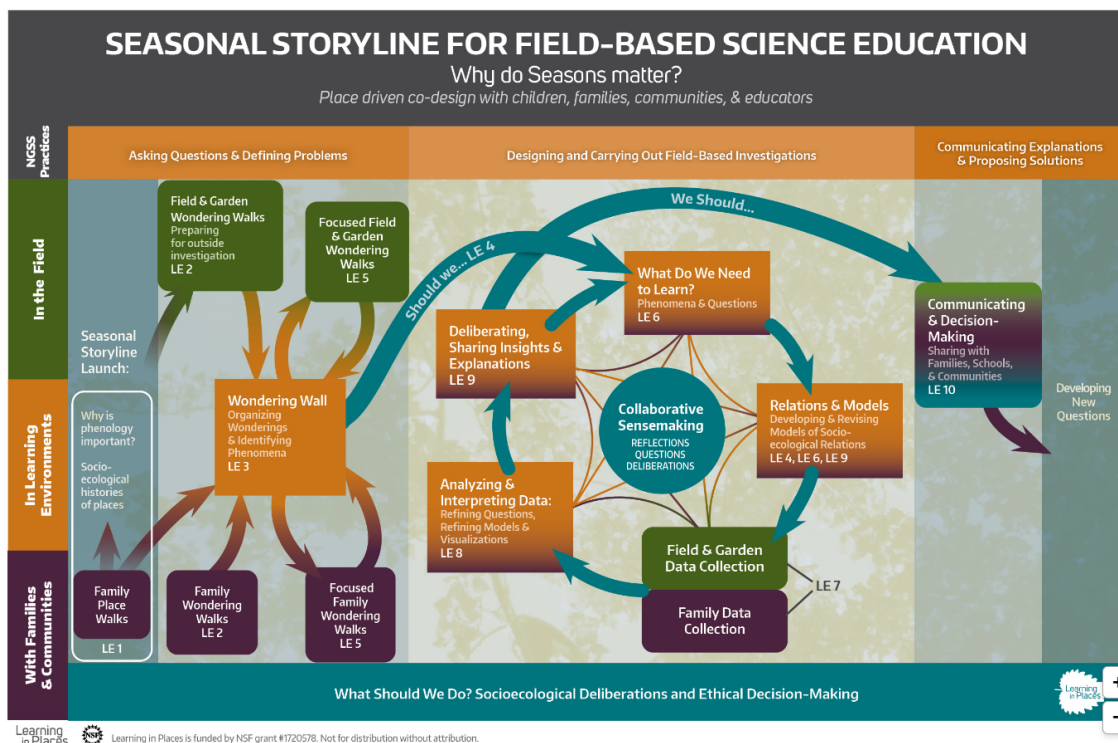


FIGURE 4-3 The Learning in Places storyline.  
SOURCE: Learning in Places Collaborative. 2023. *Our classroom storyline*. <http://learninginplaces.org/about>.

## *Curriculum and Assessment*

Teachers must have a deep understanding of their curriculum to effectively implement assessment in the classroom and build on students' cultural and social backgrounds in the process (Giamellaro et al., 2017; Ruiz-Primo, 2016). To appropriately focus the assessment of students' learning, teachers need to know where they want to take the students and how—in other words, they need to understand the curriculum deeply. A deep understanding of the curriculum includes *what* is to be learned—understanding the ambitious learning goals; *why* students should learn it—understanding the importance of current learning for future learning or for making connections with what was learned before; and *how* they will learn it—the manner in which the instructional activities and experiences will support students in meeting learning goals (Ruiz-Primo, 2016).

A deep understanding of what will be taught helps teachers determine how to gather information, what evidence will show that learning is taking place, what they need to pay attention to—what to notice—and when to gather information using formal tools. This understanding also allows teachers to design and/or select assessment tasks to provide evidence that students are learning and determine critical junctures at which to implement formal assessment checks.

### **Instruction**

Students' opportunities to learn via curricula and assessment are mediated through the process of instruction. Research has overwhelmingly indicated that the ways teachers enact lessons and conduct assessments are consequential for students' learning (e.g., Dini et al., 2020; Furtak et al., 2016; Kang et al., 2014; Shulman, 1987). Like learners, the ways teachers participate in classroom activity systems are informed by their previous experiences, their knowledge, their values, and their multiple identities.

### *Teachers' Knowledge and Its Multiple Dimensions*

Teachers' implicit theories of learning affect how they interact with learners. They can see themselves facilitating students' construction of knowledge or see their students as recipients of information provided by teachers. Educators with a multidimensional perspective on learning understand the social nature of learning and the importance of considering emotional, cultural, and cognitive facets of learning and development (see Chapter 3 of this volume, "Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems").

Teachers with a deep understanding of how learning unfolds are analytical, creative, and more selective of instructional activities, materials (e.g., texts), and assessment tasks that help learners achieve ambitious goals (Giamellaro et al., 2017; Ruiz-Primo, 2016; Shulman, 1987). Understanding the *what*, *why*, and *how* of what needs to be taught allows teachers to sequence learning activities with (1) *increasing complexity*—providing additional concepts and skills necessary to approach the tasks required from students as they progress in their learning; (2) *increasing variety*—providing additional strategies required to approach these tasks so that students can learn how certain strategies work under what conditions; and (3) *a conceptual road map*—a clear model of the over-

all activity to help students make sense of the different elements they are working on (Collins & Kapur, 2022). By viewing themselves as facilitators of learning, teachers can continuously look for evidence of learning to assist with their own instructional decisions, to provide helpful feedback to students, or to identify opportunities for students to provide feedback to one another.

Teachers can be supported to develop awareness of how their lives and the lives of their students are shaped by experiences and factors such as race/ethnicity, social class, and gender (see Chapter 5 of this volume, “Assessment Literacy and Professional Learning”). By reflecting on their own cultural reference points, teachers can expand their interpretations of student behavior and promote myriad cultural displays of learning and social interaction, just as Ms. T opened the opportunity for Simon to explain his question to the class (Hammond, 2014). Teachers can also deliberately look for barriers that may affect students’ learning; allocate or reallocate resources to ensure that every student has what they need to succeed socially, emotionally, and intellectually; and cultivate each student’s strengths (Dugan, 2021).

Teachers’ perceptions of learners can also be biased, which can in turn influence their ability to support learners through culturally responsive approaches. In a study conducted at the end of the 1970s, 100 White preservice teachers were asked to teach a student who was behind a screen (Taylor, 1979). The preservice teachers were randomly told whether the student was White or African American. When students were identified as African American, preservice teachers provided significantly less feedback after mistakes, less positive feedback after correct responses, and significantly less coaching than for students identified as White. The study also found an interaction effect of student race and gender, as White male students received the most favorable treatment and Black male students the most unfavorable. The study clearly demonstrated that student race and gender interact with teachers’ perceptions about student ability, which in turn affects teachers’ behavior.

Ultimately, ambitious teaching practices are supported when teachers have a deep understanding of their subject matter, know how students learn it, know their students, and have clear learning goals in mind (Shulman, 1987).

### *Discourse-Rich Learning Environments*

There is well-documented research literature on the types of teaching practices that facilitate student engagement in ambitious classroom learning (e.g., Hammond, 2021; Shepard, 2021; Windschitl et al., 2018). These practices share the common feature of teachers creating space for, working with, and responding to student thinking beyond the management of student behavior. Teachers can begin to create this space by utilizing strategies like varying the organizational modes of classroom activity so that learners have opportunities to engage with their ideas—and those of their classmates—individually, in pairs, in small groups, and as a whole class. To create this space, teachers can balance these different participation structures to support learners in multiple ways. Unfortunately, this condition is not always met. A study conducted in 13 classrooms showed that teachers rarely interacted with learners working individually or in small groups, thus missing the opportunity to identify and address the individual needs of learners, whether English learners or not (Solano-Flores et al., 2024).

Teachers' everyday interactions with learners are hugely consequential for their opportunities to learn (Gipps, 1999). Therefore, teachers must create space to listen and respond to learners' thinking, and use diverse modalities and strategies such as conversations in which students' reasoning, ideas, or communication styles are made explicit and can be discussed (e.g., Duschl & Gitomer, 1997). These informal interactions can open meaningful spaces for learners to think with and respond to their peers and hold students accountable to disciplinary norms, such as making arguments (Engle & Conant, 2002). "Talk moves" are a commonly recognized approach to support student thinking in these interactions, including statements that encourage students to "say more," or to build on their own or peers' thinking (e.g., Michaels et al., 2016). Teachers can also follow up on student statements by encouraging them to cite relevant evidence, provide examples, and leverage their everyday experiences.

### *Teachers' Self-Reflection*

Scholars have encouraged teachers to engage in critical self-reflection that can surface their own identities and positionalities and how they play out during daily interactions with learners (Hammond, 2014; Randall, 2021; Randall et al., 2021). For example, when teachers frame classroom conversations to allow learners to share their thinking, teachers can inadvertently shut down students' sensemaking opportunities if they limit what counts as knowledge or what might be relevant to the conversation. Teaching in a way that considers all dimensions of learning (cognitive, cultural, social, and emotional; see Chapter 3 of this volume, "Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems") requires self-reflection, checking implicit biases, practicing awareness of where one stands, and thinking about how that stance affects a teacher's relationship with their students (Hammond, 2014). This self-reflection can involve teachers recounting their lived experiences, acknowledging their identities (particularly when teachers identify as members of dominant cultures), and reflecting on how their daily decision making can influence whose ideas are recognized and shared (e.g., Wright et al., 2019).

Self-reflective teachers "make the familiar strange" by learning about their cultural values and how these values shape what they do and expect in their classrooms (Spindler & Spindler, 1982, p. 23). Spindler and Spindler (1982), rephrasing Margaret Mead, described this process as such: "If a fish were to become anthropologist, the last thing it would discover would be the water" (p. 24). Hammond (2014) also suggested that teachers can (1) first identify the cultural frame of reference in place (e.g., what they believe about learning, their models of teaching, as well as models of how they expect students to behave in class); (2) widen their cultural aperture (e.g., challenging how they interpret other people's actions or ways of knowing solely through their cultural frames); and (3) identify the triggers that may affect interactions in the classroom or can lead to miscommunication or unintended conflicts (e.g., is there an overgeneralization of certain learners' behavior by class or race?).



## Assessment

We define assessment as a process of gathering or eliciting, analyzing, and interpreting relevant information that becomes evidence about where students are in relation to the pursuit of rich learning goals, and then using this information to make decisions. Assessment refers to many things—tools to instruments to events (see Briggs, 2022; Solano-Flores, 2016; Taylor & Nolen, 2022). The idea that assessment is not the same as measuring, testing, or grading is central to this chapter’s argument—assessment does not require a numeric scale (Briggs, 2022; Taylor & Nolen, 2022). As mentioned previously, classroom assessment may involve informal observations, classroom discussions, or formal documentation about students’ learning. From a sociocultural perspective, assessment should allow multiple opportunities for students to show what they have learned in the context of, and richly informed by, their backgrounds and lived experiences. Assessment practices that best support student learning include (Taylor & Nolen, 2022):

- ensuring understanding of the learning goals by the teacher,
- ensuring that the learning goals are understood by the students,
- using assessment tasks that demonstrate the learning goals,
- paying attention to factors that may affect learners’ performance,
- evaluating learners’ performance based on learning goals—not unrelated factors, and
- providing opportunities to close the gap between where learners are and where they need to be through feedback, revision, and promoting self-assessment.

Classroom assessment consists of a diverse set of strategies to gather information about student learning (Fine & Furtak, 2020; Furtak & Ruiz-Primo, 2008; Ruiz-Primo, 2011; Ruiz-Primo & Brookhart, 2018; Ruiz-Primo & Furtak, 2006, 2007):

1. The assessments should include a range of *informal to formal* assessment tasks (e.g., instructional dialogues, quizzes).
2. They should be *multimodal* in documenting students’ learning (e.g., performance, explanations, graphical representations).
3. They should appear in different *organizational modalities* (i.e., diverse size and composition—individual, pairs, small groups, whole class; Ruiz-Primo et al., 2016).
4. They may have *different foci* (e.g., tasks that focus on content and skills and tasks that focus on reinforcing metacognitive forms of thinking—metacognitive monitoring, metacognitive control, and self-regulation).

Classroom assessments sometimes involve grading—but not always. While grading is a difficult practice, it is necessary in the current education system to document academic achievement (Shepard, 2019). In the context of classroom assessment, it is important to distinguish between assessment intended to assist student learning—which requires no grading since its purpose is not to measure but to support learners—and assessment of individual achievement, which involves summative assessment along

with other pieces of evidence reflected in a grade. Grading should be handled with care because it can affect students' motivation, self-confidence, and efficacy, as well as unearth problems that are not immediately apparent. Grading can promote motivation to achieve good grades, rather than a motivation to learn well, to have deep learning. When grading is used, comparisons with peers are inevitable. Another issue is that in many cases grading does not focus only on students' learning, but also on other factors that may not necessarily reflect what students know and can do. For example, teachers tend to include student effort in the classroom in their grading practices, which is different from what students know and can do (Brookhart, 2013; Brookhart et al., 2016). Omnibus grading—grading that involves factors other than learning—is an inappropriate assessment practice; it is more suitable to bias (Feldman, 2019).

Students are a crucial element of classroom assessment and should be engaged with as such. Two strategies that promote student engagement in assessment practices are self- and peer assessment (judging their own work or the work of others) (Formative Assessment for Students and Teachers State Collaborative on Assessment and Student Standards, 2018; Leahy et al., 2005; Wiliam & Thompson, 2008). Self- and peer assessment are both important practices that can support metacognitive awareness and self-regulation. For these practices to work, students need to understand what is expected—they need to be provided with criteria that help them monitor their learning. Defining the criteria by which students assess their work is critical, as is helping students to decide what to do next (e.g., ask for help to the teacher or peers; re-read the information). When students examine the work of their peers, knowing that their peers will also look at their work, it helps them to develop internal standards to evaluate their work, thus improving self-regulation (Bourgeois, 2016; Panadero et al., 2016).

Classroom assessment is more effective when learning goals are clear, students know the criteria for success, and there are opportunities to provide and use feedback. Effective feedback is based on learning goals and success criteria—or “student look-fors” (Moss & Brookhart, 2012). Students should know what they are supposed to be learning and the criteria by which their learning will be assessed. Research has shown that students who understand success criteria produce better work and are more self-regulated learners; teachers who provide clear success criteria use students' work and responses more efficiently as evidence to support their instructional decisions; and teachers use insights gleaned from the classroom to provide helpful feedback (Kroog et al., 2016; Minstrell et al., 2009; Panadero et al., 2012; Ruiz-Primo & Kroog, 2018; Ruiz-Primo et al., 2014a, 2016).

There are many ways to characterize feedback, but an important distinction should be made between *evaluative* and *descriptive* comments (Ruiz-Primo & Brookhart, 2018). *Evaluative* feedback (e.g., your response is incorrect; good work!) does not help to improve students' learning or develop strategies to approach problems—it can create more “noise” than true change and should be avoided (Hattie & Timperley, 2007). In contrast, *descriptive* feedback (e.g., your response is missing X, it would be important for you to always check Y) helps students to focus on their learning and understand where they are and what they can do to move forward. Statements that are clear and useful contain information to influence students' future performance. *Descriptive* feedback has the potential to shape the student's motivation toward achieving learning goals (Dweck, 1986). *Descriptive* feedback guides the student's attention to the process

underlying the task (e.g., “you need to remember the three characteristics that define X”) or the students’ product (e.g., “the table is incomplete, it is missing X and Y”), rather than characteristics of the student (e.g., you are so smart!). *Descriptive* feedback is process oriented, it focuses on aspects of the student’s performance, including describing what the student has accomplished and/or what needs to be worked on or improved; clarifying the process the student needed to engage in to do the task; and/or helping the student compare previous and current achievements or performance. High-level feedback involves the student as a learner who can reflect on his or her learning, helping the student to make connections about what has been learned at any given point (Ruiz-Primo & Brookhart, 2018).

There are two generally accepted forms of classroom assessment: formative assessment, or assessment for learning, which is used to draw out what students know and can do while learning is in progress; and summative assessment, or assessment of learning, which takes place at the end of learning experiences to certify individual achievement—most likely for grading purposes (Shepard, 2019). For any type of classroom assessment, summative or formative, assessments should be closely aligned with and connected to the curriculum that students are experiencing. When curriculum, instruction, and assessment support each other, there is horizontal coherence in the classroom activity system (National Research Council, 2001; see Chapter 2 of this volume, “The Struggle to Implement Balanced Assessment Systems: Explanations and Opportunities”; Shepard et al., 2018a). For formative assessments, the grain size of the assessments is small, meaning specific aspects of the disciplinary knowledge and practices used can be interrogated (e.g., specific concepts and practices that are part of the big ideas being pursued). External assessments, such as state assessments, cannot focus on these specific aspects because the grain size is larger (e.g., based on standards). Therefore, classroom assessments should not be separated from curriculum, instruction, and student opportunities to learn (Gee, 2003; Stobart, 2005). Classroom assessment should reflect a clear alignment between the “what” and “how” of instruction and the “what” and “how” of the assessment strategies used (see Chapter 3 of this volume, “Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems”).

### *Formative Assessment*

*Formative assessment* is an ongoing process involving planned and unplanned events that provide opportunities to gather or elicit information that becomes evidence about student learning. This evidence can be used by teachers and students to support students in pursuing rich disciplinary learning goals and supporting students in becoming self-directed learners (e.g., *Formative Assessment for Students and Teachers State Collaborative on Assessment and Student Standards*, 2018; Ruiz-Primo, 2010; Ruiz-Primo & Furtak, 2006, 2007). Formative assessment enables teachers and students to support students’ future learning, either by adjusting instruction to achieve learning goals or by providing focused feedback to support their learning (Black & Wiliam, 1998; Sadler, 1989). Formative assessment focuses on continuously gathering information to allow immediate action to support student learning and instruction in pursuit of daily learning goals.

Four critical formative assessment practices and activities include (Herman, 2016; Ruiz-Primo, 2010; Ruiz-Primo et al., 2015)

1. Sharing and clarifying learning goals, expectations, and/or quality/success criteria;
2. Gathering or eliciting evidence of where learners are in achieving learning goals;
3. Analyzing and interpreting the information collected; and
4. Acting on that information to improve students' learning through instructional adjustments or focused feedback.

Formative assessment can be conceptualized as a continuum from very informal (students' questions, which cannot be predicted) to very formal, requiring specific documentation of students' performance (e.g., quizzes). Considering formative assessment as a continuum presents many possibilities to gather and respond to information as needed. Gathering informal information about what students know and can do often happens "on-the-fly" in classroom discourse (Jordan & Putz, 2003; Ruiz-Primo, 2011; Ruiz-Primo & Furtak, 2006, 2007). For example, a question asked by a student—like Simon's question from the vignette that opens this chapter—provides an opportunity for a teacher to assess the student (e.g., what information can be gathered about this student's learning from this question?). It also provides an opportunity to analyze and interpret the student's question within the frame of the learning goals (e.g., Ms. T needed to quickly navigate the intellectual substance of Simon's question before acting with a key instructional move). A student's question provides an opportunity to take action with an *instructional move* (like Ms. T asking Simon to explain his question and using Simon's response to move forward the discussion), or *useful feedback* that is focused and explicit in moving learning forward.

Another type of formative assessment is when a teacher asks specific, carefully crafted, and planned questions to the students. These "back-pocket" questions are challenging and require explanations, justifications, and analysis, inviting students to think critically about their ideas (Windschitl et al., 2018). These types of questions are also assessment tasks that will provide evidence about students' learning. The responses will require analysis and interpretation "on-the-fly," as well as certain responsive actions (e.g., an instructional move or feedback) that are only possible when learning goals are clear.

Yet another type of formative assessment is when teachers provide tasks for all students to gather information from each of them. The teacher can use diverse sources—handouts or exit tickets (that do not look like tests or assessment tasks), quizzes (more test-like tasks), and everything in between (e.g., classwork or homework)—to gather information from students in a formal way. Once this information is collected, teachers should use their understanding of the discipline at hand to analyze and interpret student responses before taking action based on that interpretation.

In our observational studies, we have tracked classroom discussions that show this pattern of eliciting, interpreting, and acting on information to inform teachers' instructional moves (Furtak et al., 2017; Ruiz-Primo & Furtak, 2006, 2007). In each of these cases, teacher and student participation in assessment practices were organized

by sets of tasks or tools to gather information from all students at the same time—the information was then interpreted to assist student learning toward the learning goal by adapting instruction or providing feedback (Furtak et al., 2019).

Opportunities to provide descriptive feedback to students are critical in formative assessment. Given the many opportunities provided to gather information about students' learning in formative assessment, teachers can respond almost immediately with short targeted comments or with an instructional move to help students reflect on their thinking or performance (e.g., move students around in different small groups; Ruiz-Primo et al., 2014a, 2015).

Most of the activities conducted in classrooms are intended to introduce new topics or serve as purposeful practice (e.g., homework) that can improve students' learning based on feedback. These activities, in a coherent and appropriate combination, provide evidence of how students move toward expertise. At the same time, teachers expect that students will make mistakes as part of the learning process. These mistakes are not only expected, but desirable because they provide the opportunity to improve learning (Feldman, 2019; McMillan, 2018; Wiliam, 2018). Grades provided to in-progress learning activities are not appropriate because they may negatively impact the learners' motivation and, therefore, their learning. Grading correct responses in activities meant to encourage taking risks and making mistakes can send students the wrong signal and reduce their academic confidence (Brookhart, 2013; Feldman, 2019). Attention, organizational skills, and collaboration are appropriate targets for formative assessment, but not for grading (Shepard, 2019). Grades should be based only on appropriate evidence of what students know and what they can do as participants within a discipline.

### *Summative Assessment*

The most accurate information about what students have learned is their performance at the end of the learning process (Feldman, 2019). Summative assessments are formal tasks (e.g., a test or a final project) used to gather evidence about students' learning or mastery of the knowledge, skills, and practices that were the instructional focus across a specific period. Since the summative test (administered at the end of an instructional unit) or summative formal tasks (an end of semester project) are usually used for grading purposes, teachers should seek to maximize their students' performance by ensuring that all students are fairly assessed, have opportunities to demonstrate what they have learned, and that the evidence collected from the assessment is as free as possible from extraneous influences (e.g., use of certain language that may interfere with what is being assessed). When summative assessment tasks are being selected, teachers should ensure that the tasks are aligned with the opportunities students had to learn the material (e.g., does the assessment map to what was taught in the classroom?) and that the characteristics of the task do not favor a subgroup of students (e.g., the assessment does not include features that prevent some groups of learners—such as multilingual learners or students with disabilities—from demonstrating what they have learned).

The design and development of quality summative assessment tasks matter and should provide opportunities for all students to demonstrate what they have learned. It is important to allow learners to provide evidence of what they know and therefore



to identify and utilize tasks that can accurately reflect what they have learned. An end-of-year grade will include different sources of information that are considered important evidence of students' learning. Grading should be based on rich tasks (e.g., project-based tasks) that best reflect the learning being pursued.

Rich tasks represent as fully as possible the ambitious learning goals set for the learners. They are not "more of the same" of what learners experienced during the instruction, but instead challenge learners to use what they have learned in different ways, to promote knowledge that is more conditional, strategic, and simulates the type of thinking required "in the real world." Tasks that support ambitious learning and teaching are also cognitively demanding, requiring students to go beyond factual recall and engage in disciplinary practices (e.g., Tekkumru-Kisa et al., 2020). These tasks also place learning in contexts that students are likely to encounter in their everyday lives. For example, learners might be asked to conduct a short- or long-term investigation in which the applications of what they have learned are not straightforward, but require examining documents, revising what they have done and adapting it to the new situation, or developing arguments that support the decisions they made to conduct the investigation.

To inform parents about their children's progress, teachers can provide examples of their performance—what they know and can do—at different times during the year, rather than only sharing grades. These examples can reflect critical milestones (Shepard, 2019; Shepard et al., 2018a, 2018b). Practices that consider collaboration or effort in the provision of grades make the grading less accurate, encourage performance orientation, and decrease intrinsic motivation. Instead, motivation can be built by offering relevant tasks to the students—when possible, a choice of tasks—and focusing both on academic success and opportunities to improve.

### **Classroom Learning Culture**

The previous elements of the activity system all exist within a social and physical context called a classroom learning environment or learning culture. Learning cultures are developed and designed with specific characteristics and organizational structures that create communities of practice with shared norms, routines, values, practices, discourse patterns, and particular physical, symbolic, and non-verbal cues (e.g., artifacts, images, icons; Gay, 2018) that support learners.

Classroom learning cultures create contexts that are suitable for learning and have an impact on learners' behavior and intellectual functioning. When classroom environments are safe and engaging and learning is supported and rewarded, students are connected to the classroom community and feel efficacious (Hammond, 2021; Melnick et al., 2017; Shepard, 2021). This type of learning environment allows students "to develop the social and emotional, as well as academic skills, habits, and mindsets needed to succeed in life" (Melnick et al., 2017, p. v).

Classroom cultures are initiated at the beginning of the school year and can be cultivated over time. The importance of the first days of the school year, and even the first hours of the first day of a class are critical for helping students understand the learning expectations, rules, and norms that will be followed, as well as orienting students toward overarching learning goals (Boaler, 2022; Jackson, 1971; Seidel et al.,

2005). Learning goals support students' experiences and who they are as members of the learning community.

In a culturally and socially responsive classroom learning culture, teachers select or design rich and authentic tasks based on students' interests and experiences, and facilitate classroom discourse to help students develop disciplinary knowledge relevant to their lives. Assessment practices should provide students with opportunities to show their learning using tools that consider their cultural and social identities, backgrounds, and experiences (Taylor & Nolen, 2022).

Culturally responsive classrooms provide students with participation opportunities that invite their "funds of knowledge" into learning activities. Moll and Greenberg (1990) described a sixth-grade classroom that included a high percentage of Mexican students who had problems with writing. To help students actively create and shape their writing tasks, the teacher asked the students to brainstorm a list of writing topics. They discussed the topics that were of main interest and then chose construction and building—one of the most prominent direct experiences for these learners outside the classroom and a clear example of the use of "funds of knowledge" (Moll & Greenberg, 1990; Moll et al., 1992). Learners were asked to design models of houses or buildings and then write about how they were planning to build the model with paper and other materials. The teacher then introduced the idea of learners researching design and construction. The students visited the library and obtained information about different ways of building structures. Learners then wrote brief essays about construction—some students wrote about how to build a model, others created stories about construction, and yet others compared construction to the human body. Learners found writing more exciting by focusing on something interesting to them (Moll & Greenberg, 1990; Moll et al., 1992).

These learning activities grew out of the teacher's initial focus on the students' interest in writing. The teacher provided feedback to the students' writings as they wrote about their construction stories. Over time, learners' writing showed better punctuation, fewer orthographic errors, and improved narrative structures. The students' enthusiasm about construction led the teacher to suggest inviting their parents as experts (e.g., constructors, plumbers, carpenters) to provide more information about the topic—what Moll and Greenberg (1990) called "mobilizing funds of knowledge." The parents and relatives of the learners made an intellectual contribution to the classroom by sharing their knowledge and experience (e.g., how to estimate materials or measure areas).

This example shows how learners can increase their curiosity and disciplinary engagement when their experiences, interests, and identities are centered in classroom learning environments. This example also illustrates a distributed view of knowledge construction with a cultural orientation: learners' interests and cultural backgrounds were considered in critical instructional decisions and the teacher welcomed their contributions by considering their strengths (e.g., proposing questions in Spanish). The teacher helped the learners actively create and shape the instructional activities by connecting the activities to students' "funds of knowledge." Formal and informal assessments included the learners' brief essays and discussions (e.g., how to formulate appropriate questions) as a source of information about how to support the students in future learning.

### *Norms, Routines, and Tools in Culturally Responsive Learning Environments*

The culturally responsive classrooms described in this chapter are deliberately created around norms and routines that support student participation. They have organizational structures that create specific classroom communities of practice with shared norms, routines, values, discourse patterns, and particular physical and symbolic cues. Culturally responsive classrooms are designed to be safe spaces for learners, where they feel a shared sense of belonging and respect (Melnick et al., 2017; Shepard, 2021). Norms (rules and expectations), routines (instructional segments that occur regularly), and tools (artifacts that promote certain forms of interactions) all contribute to classroom culture.

Some norms, routines, and tools are more conducive to gathering information about students' learning (e.g., discussions); supporting learners' engagement and contributions; and allowing for meaningful collaboration among learners of different racial identities, economic circumstances, cultures, and ability levels (Kahlenberg et al., 2019; Killen et al., 2015; Ruiz-Primo et al., 2022). They can provide students with *affordances* ("action possibilities"; Gee, 2008, p. 81) or *constraints* (disengagement). Helping students to see these possibilities allows them to transform the *affordances* into effective actions that they can take advantage of. Peers are part of the learning environment that also offer diverse *affordances* through their knowledge and skills (Gee, 2008).

Identifying the norms that are conducive to learners' engagement, agency, and productive discursive practices—and therefore, formative assessment—is critical. For example, a teacher may establish the norm that students should draw on their own experiences to make sense of what they are learning in school, and as such, everyday examples are welcome (e.g., Odden & Russ, 2018). Teachers could also emphasize a norm that errors are welcome because they are a very important part of how we all learn and change.

Some classroom routines are more conducive than others to the implementation of assessments with a formative purpose. For example, the "Navigating Routine" in the OpenSciEd storyline in Figure 4-2 invites students to revisit the driving question for a unit, determine what they have learned so far, and determine what to investigate next as a classroom community (OpenSciEd, 2022). In another example, we observed a teacher establishing "homework circles" on the first day of the school year, a routine that allows learners to talk to each other about the problems they encounter in completing their homework, providing an opportunity for learners to support each other (Ruiz-Primo et al., 2022).

Classroom tools can also be used to support learners. As an example, a teacher hung science concepts from the ceiling in different languages on double-sided pieces of cardboard (Ruiz-Primo et al., 2014b). During the unit, when introducing a new concept, the teacher pointed out the words attached to the ceiling to help students to recognize the term in English and in other languages. We also observed a teacher in a multilingual classroom use her hands to explain to students what they needed to do—to draw a scatterplot. The teacher moved her hands up and down and right to left to represent the two coordinates and use her right hand to exemplify how students needed to plot the data using the coordinates. She explained with her hands the changes in the value of the dependent variable and she modulated the tone of her voice while describing the

relationship (Ruiz-Primo et al., 2014b). These and other semiotic modalities (symbolic, physical, or verbal) can provide affordances that help multilingual students participate more completely in classroom discussions—for example, learners using different vocal intonations to support their description of different sounds made when guitar strings are plucked (Suárez, 2020).

Learners can be involved in developing norms and routines in the classroom. Teachers can use different practices in the first days of the school year, including providing scenarios in which norms or routines can be applied and then asking learners questions about what they should do in the scenario (Ruiz-Primo et al., 2022). Involving learners in the development of these types of cultural practices can make them feel like they are contributing to the characteristics of their classroom community.

The cultural context can also support assessment practices and cultural responsiveness. How students participate, how they feel about making mistakes, and how much they value cultural differences will depend on how the characteristics of the learning environment are established. Creating caring learning communities that use cultural knowledge of ethnically diverse cultures to guide the curriculum, instruction, and assessment and where differences are valued creates a safe climate for learning. Culturally responsive learning environments value the traditions, languages, and communication styles of the students to create a community of practice. Learning environments that create communities of practice lead to a sense of “ownership characterized by personal investment and mutual dependency” (Collins & Kapur, 2022, p. 163).

### **BRINGING IT TOGETHER: IMPLICATIONS FOR THE DESIGN OF CLASSROOM ASSESSMENT TO SUPPORT AMBITIOUS INSTRUCTION**

To illustrate how a classroom activity system lens helps with understanding how assessment can support ambitious teaching, we return to Ms. T’s classroom, first shown at the beginning of this chapter. Ms. T’s learning environment was designed around the topic of growing pumpkins, a plant that students have prior experience with and that they can directly observe in their classroom. The classroom conversation that led to Simon’s question—and Ms. T’s response—is embedded in a larger curriculum that invites everyday examples and language to help learners understand the life cycles of plants. Across the course of a larger unit, Ms. T used a variety of classroom participation structures and resources to support students in sharing their thinking. She employed multiple talk moves that drew out student thinking and encouraged students to provide further elaboration to support their ideas. The task was completed in a classroom culture in which students knew their ideas were valued and where they felt safe sharing what they know with each other and their teacher. The features of the classroom activity system described in the previous sections are summarized in Table 4-1.

#### **Culturally and Socially Relevant Assessments**

Classroom assessments need to be designed to be more culturally and socially relevant for diverse students. This involves students, parents, and community members—not just teachers—contributing to and even participating in the assessment design process (Taylor & Nolen, 2022). Culturally and socially relevant assessments, like any

**TABLE 4-1** Characteristics of Classroom Activity Systems That Support Ambitious Instruction and Assessment by Element

Element	Classroom Assessment Systems Should
Learners	<ul style="list-style-type: none"> <li>• Be centered on learners’ interests and identities</li> <li>• Draw on learners’ linguistic and cultural capital</li> <li>• Nurture the development of learners’ metacognition and self-regulation</li> </ul>
Curriculum	<ul style="list-style-type: none"> <li>• Be designed to respond to and sustain learners’ knowledge and practice</li> <li>• Be structured around learners’ understanding and explicit learning goals</li> <li>• Be designed to support teachers’ understanding of what is to be learned, why it is being taught, and how it will be taught</li> <li>• Be designed to reflect increasing complexity and variety</li> <li>• Be designed to support learners’ organization and representation of information</li> </ul>
Instruction	<ul style="list-style-type: none"> <li>• Be built on teachers’ knowledge of the discipline and curriculum</li> <li>• Be based on teachers’ conception of learning</li> <li>• Be based on a discourse-rich learning environment</li> <li>• Be informed by self-reflection</li> </ul>
Assessment	<ul style="list-style-type: none"> <li>• Focus on a good understanding of the learning goals</li> <li>• Use both formative and summative assessments</li> <li>• Align to everyday learning</li> <li>• Align to classroom practices that draw out and work with student thinking</li> <li>• Provide multiple self- and peer assessment opportunities</li> <li>• Be comprehensive, multimodal, formal and informal, and cognitively challenging</li> </ul>
Classroom Learning Culture	<ul style="list-style-type: none"> <li>• Be culturally responsive</li> <li>• Include norms, routines, and tools that support student participation</li> <li>• Be safe and engaging</li> <li>• Support and recognize learning</li> <li>• Connect students to a classroom community</li> <li>• Support the development of social, emotional, and academic skills, habits, and mindsets needed to succeed in life</li> </ul>

assessments, should consider certain characteristics to support appropriate interpretations of student performance, such as fairness and representation (Gee, 2003; Taylor & Nolen, 2022) and cognitive demands (Tekkumru-Kisa & Stein, 2015).

There are a range of emerging approaches to this type of co-designing with multiple stakeholders, including student interest surveys, which solicit input from students that teachers can use to design or adapt classroom assessments like those in use by Open-SciEd, or ongoing and in-depth collaboration with communities and families (Earnest et al., 2023; Edelson et al., 2021; Tzou et al., 2021).

### Fairness and Representation

In the classroom context, fairness is a sociocultural issue rather than a technical one (Gee, 2003; Stobart, 2005). There are critical questions to ask about fairness in the testing, assessing, and measuring context (adapted from Stobart, 2005):



- What knowledge is assessed and equated with learning?
- Are the forms, content, and mode of an assessment appropriate for different groups and individuals?
- Is the range of cultural knowledge and practices reflected in definitions of learning?
- How does cultural knowledge mediate responses to assessments in ways that alter what is being assessed?

Taylor and Nolen (2008) suggest asking students to generate ideas to be assessed based on their learning experiences (e.g., instructional activities, classroom discussions). It is important to also question: (1) what content is to be taught, learned, and assessed (Taylor & Nolen, 2022); (2) who has the power to determine what content is to be taught (Randall, 2021); and (3) who is being considered in gathering the full range of understandings in the classroom? (Ruiz-Primo & Brookhart, 2018; Solano-Flores, 2016).

There are many considerations for the design of classroom assessments that will better represent all that learners know and can do. We have argued that centering the interests and identities of marginalized learners in the classroom can reposition expertise and help to highlight the assets and resources that students bring to their learning. For example, Kang and colleagues (2022) illustrated how different forms of classroom assessment tasks can broaden what is known about learners. In a high school physics unit, learners completed different forms of tasks aligned with next-generation standards, including a claim-evidence-reasoning assessment and a letter to a loved one. Both assessments were intended to create expansive space for learners to show what they know about how different car designs help to make them safe. However, for some learners, what they knew was better captured in the letter—which many wrote in their home language, rather than English—explaining how they would design a dream car. Many additional designs can be integrated into tasks to make them applicable to all learners, such as decreasing language load, including pictures and other visual representations, and breaking longer tasks into smaller pieces that can be done in different sequences (Fine & Furtak, 2020).

### **Cognitive Demands**

Assessment tasks that support ambitious instruction go beyond simple factual recall or assessing knowledge or practices in isolation and are designed to move students from their comfort zone into a learning zone where the level of struggle is well calculated—not too easy and not too hard. There should be tasks that challenge the students outside of their comfort zone but should not be so difficult that the tasks demotivate the learners. This type of task seeks to elicit students' integrated knowledge and practice while they are doing disciplinary tasks (Tekkumru-Kisa & Stein, 2015). The design of the tasks signal what is important and how students can show their intellectual activity and engagement in disciplinary practices (Tekkumru-Kisa et al., 2015). To support student learning, assessment tasks need to be designed with a purpose in mind.

## CONCLUSION

In this chapter, we have articulated elements of ambitious instruction; their relation to ambitious classroom assessment; and the importance of taking a broader lens to encompass learners, curriculum, instruction, and classroom learning cultures as well as assessment. These concepts help to illuminate multiple aspects of classroom learning and teaching that are essential to realizing assessments that are cognitively demanding, build on students' prior experiences, and support their engagement in disciplinary practices.

While we have deliberately focused on classrooms, we acknowledge that there are other influences on these systems that have not been discussed in this chapter. School, district, and state policies and initiatives; district and interim assessments; and accountability measures enforced by statewide tests can also enable—or constrain—what is possible in classrooms (e.g., Au, 2007). These outside-the-classroom influences will be discussed in greater depth later in this volume (e.g., 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”).

This chapter does not include all research on learning, instruction, and assessment. Rather, it focuses on what we consider to be important information to think about for each element of the system. We hope that the information provided in this chapter can help to broaden the lens when considering assessment within larger classroom activity systems and inspire readers to look deeper into each element.

We acknowledge that approaches to assessment systems that interrogate the constructs being assessed and whose values are represented is an emerging area of research that needs more study (Randall, 2021). To realize visions of ambitious instruction that broaden access and opportunity for all learners, we must continue to consider how all elements of an activity system can support equity and justice (Kang & Furtak, 2021).

As we look to a future in which classroom assessment is embedded within ambitious approaches to classroom teaching, we emphasize the benefits shown by collaborative efforts at scale. The approach of long-term, mutually beneficial partnerships centered on problems of practice emerging from educational organizations can support the design of assessment systems at scale by building infrastructures to support professional learning, curriculum design, and assessment (see Chapter 5 of this volume, “Assessment Literacy and Professional Learning”; Chapter 6 of this volume, “District and School Practices and Assessments to Support a Learning-Centered Vision”; Penuel et al., 2011). Furthermore, groups of teachers seeking to learn new practices can be connected across schools to support the spread of local innovation (Thompson et al., 2019).

Gathering information from different sources to develop a more accurate vision of students' learning—Cronbach's (1990) view of what constitutes assessment—should be done at the classroom level, where the information is more likely to directly impact instructional practices. To support students' learning, formative assessment should be aligned with ambitious instruction and attention should be paid to teacher preparation programs and professional development. Ambitious teaching that provides more and better opportunities for students to explain their thinking and reasoning and to develop disciplinary practices requires a deep understanding of the subject matter and sociocultural practices that allow students to be part of a community of cognitive

apprenticeship. There is no question that proper alignment of classroom assessment practices with ambitious instruction requires changes at higher levels of the education system (e.g., modifying grading practices).

Assessment is a powerful tool that can serve stakeholders at all levels of the educational system if they are committed to improving their understanding of the role that assessment has at different levels, its purposes and uses, and the characteristics of its design and development (Ruiz-Primo et al., 2024). If assessment's role at different levels is better understood, it can be properly designed and used to support student learning.

## REFERENCES

- Alexander, P. A. (2006). *Psychology in learning and instruction*. Pearson, Merrill, Prentice Hall.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258–267. <https://doi.org/10.3102/0013189X07306523>.
- Au, K. H-P., & Jordan, C. (1981). Teaching reading to Hawaiian children: Finding a culturally appropriate solution. In H. T. Trueba, G. P. Guthrie, & K. H-P. Au (Eds.), *Culture and the bilingual classroom: Studies in classroom ethnography* (pp. 139–152). Newbury House Publishers, Inc.
- Bang, M., & Marin, A. (2015). Nature-culture constructs in science learning: Human/non-human agency and intentionality. *Journal of Research in Science Teaching*, 52(4), 530–544. <https://doi.org/10.1002/tea.21204>.
- Bang, M., & Medin, D. (2010). Cultural processes in science education: Supporting the navigation of multiple epistemologies. *Science Education*, 94(6), 1008–1026. <https://doi.org/10.1002/sce.20392>.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>.
- Boaler, J. (2022). *Mathematical mindsets: Unleashing students' potential through creative mathematics, inspiring messages, and innovative teaching* (2nd ed.). Jossey-Bass.
- Bourgeois, L. (2016). Supporting students' learning: From teacher regulation to co-regulation. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 345–363). Springer. [https://doi.org/10.1007/978-3-319-39211-0\\_20](https://doi.org/10.1007/978-3-319-39211-0_20).
- Briggs, D. C. (2022). *Historical and conceptual foundations of measurement in the human sciences: Credos and controversies*. Routledge: Taylor & Francis Group.
- Brookhart, S. M. (2013). Grading. In J. H. McMillan (Ed.), *SAGE Handbook of Research on Classroom Assessment* (pp. 257–271). Sage.
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803–848. <https://doi.org/10.3102/0034654316672069>.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 6(1), 32–42. <https://www.jstor.org/stable/1176008>.
- Bruner, J. (1996). *The culture of education*. Harvard University Press.
- Choppin, J., Davis, J., Roth McDuffie, A., & Drake, C. (2021). Influence of features of curriculum materials on the planned curriculum. *ZDM – Mathematics Education*, 53, 1249–1263. <https://doi.org/10.1007/s11858-021-01305-7>.
- Collins, A., & Kapur, M. (2022). Cognitive apprenticeship. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (3rd ed., pp. 156–174). Cambridge University Press.
- Conley, D. T. (2018). *The promise and practice of next generation assessment*. Harvard Education Press.
- Council of Chief State School Officers. (2018). *Formative Assessment for Students and Teachers (FAST) State Collaborative on Assessment and Students Standards (SCASS)*. <https://ccsso.org/resource-library/fundamental-insights-about-formative-assessment>.
- Cowie, B., Jones, A., & Otrell-Cass, K. (2011). Re-engaging students in science: Issues of assessment, funds of knowledge and sites for learning. *International Journal of Science and Mathematics Education*, 9(2), 347–366. <https://doi.org/10.1007/s10763-010-9229-0>.
- Cronbach, L. J. (1990). *Essentials of psychological testing* (5th ed.). Harper Collins Publishers.

- deBarger, A. H., Penuel, W. R., Moorthy, S., Beauvineau, Y., Kennedy, C. A., & Boscardin, C. K. (2017). Investigating purposeful science curriculum adaptation as a strategy to improve teaching and learning. *Science Education*, *101*(1), 66–98. <https://doi.org/10.1002/sce.21249>.
- Dini, V., Sevian, H., Caushi, K., & Orduña Picón, R. (2020). Characterizing the formative assessment enactment of experienced science teachers. *Science Education*, *104*(2), 290–325. <https://doi.org/10.1002/sce.21559>.
- Dugan, J. (2021). No shortcut: Avoiding equity traps and tropes. In S. Safir & J. Dugan (Eds.), *Street data: A next-generation model for equity, pedagogy, and school transformation* (pp. 25–45). Corwin: A Sage Company.
- Duschl, R. A., & Gitomer, D. H. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment*, *4*(1), 37–73.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*(10), 1040–1048. <https://doi.org/10.1037/0003-066X.41.10.1040>.
- Earnest, D., Randall, J., Sweeney, S., & Thota, N. (2023, April 13–16). *Investigating first-grades' informal computational thinking: A fund of knowledge approach* [Paper presentation]. AERA 2023 Annual Meeting, Chicago, Illinois.
- Edelson, D. C., Reiser, B. J., McNeill, K. L., Mohan, A., Novak, M., Mohan, L., Affolter, R., McGill, T. A. W., Buck Bracey, Z. E., Deutch Noll, J., Kowalski, S. M., Novak, D., Lo, A. S., Landel, C., Krumm, A., Penuel, W. R., Van Horne, K., González-Howard, M., & Suárez, E. (2021). Developing research-based instructional materials to support large-scale transformation of science teaching and learning: The approach of the OpenSciEd Middle School Program. *Journal of Science Teacher Education*, *32*(7), 780–804. <https://doi.org/10.1080/1046560X.2021.1877457>.
- Engeström, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work*, *14*(1), 133–156. <https://doi.org/10.1080/13639080020028747>.
- Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction*, *20*(4), 399–483. [https://doi.org/10.1207/S1532690XCI2004\\_1](https://doi.org/10.1207/S1532690XCI2004_1).
- Erickson, F., & Mohatt, G. (1982). Cultural organization of participation structures in two classrooms of Indian students. In G. Spindler (Ed.), *Doing the ethnography of schooling: Educational anthropology in action* (pp. 132–174). Waveland Press, Inc.
- Feldman, J. (2019). *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms*. Corwin: A Sage Company.
- Fine, C. G. McC., & Furtak, E. M. (2020). A framework for science classroom assessment task design for emergent bilingual learners. *Science Education*, *104*(3), 393–420. <https://doi.org/10.1002/sce.21565>.
- Flores, N., & Rosa, J. (2015). Undoing appropriateness: Raciolinguistic ideologies and language diversity in education. *Harvard Educational Review*, *85*(2), 149–171. <https://doi.org/10.17763/0017-8055.85.2.149>.
- Formative Assessment for Students and Teachers State Collaborative on Assessment and Student Standards. (2018). *Revising the definition of formative assessment*. <https://ccsso.org/resource-library/revising-definition-formative-assessment>.
- Furtak, E. M., & Ruiz-Primo, M. A. (2008). Making students' thinking explicit in writing and discussion: An analysis of formative assessment prompts. *Science Education*, *92*(5), 799–824. <https://doi.org/10.1002/sce.20270>.
- Furtak, E. M., Kiemer, K., Circi, R. K., Swanson, R., de León, V., Morrison, D., & Heredia, S. C. (2016). Teachers' formative assessment abilities and their relationship to student learning: Findings from a four-year intervention study. *Instructional Science*, *44*, 267–291. <https://doi.org/10.1007/s11251-016-9371-3>.
- Furtak, E. M., Ruiz-Primo, M. A., & Bakeman, R. (2017). Exploring the utility of sequential analysis in studying informal formative assessment practices. *Educational Measurement: Issues and Practices*, *36*(1), 28–38. <https://doi.org/10.1111/emip.12143>.
- Furtak, E. M., Heredia, S. C., & Morrison, D. (2019). Formative assessment in science education: Mapping a shifting terrain. In H. L. Andrade, R. E. Bennett, & G. J. Cizek (Eds.), *Handbook of Formative Assessment in the Disciplines* (pp. 97–125). Taylor & Francis Group. <https://doi.org/10.4324/9781315166933>.
- Gay, G. (2018). *Culturally responsive teaching. Theory, research, and practice* (3rd ed.). Teachers College Press.
- Gee, J. P. (2003). Opportunity to learn: A language-based perspective on assessment. *Assessment in Education: Principles, Policy & Practice*, *10*(1), 27–46. <https://doi.org/10.1080/09695940301696>.



- Gee, J. P. (2008). A sociocultural perspective on opportunity to learn. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 76–108). Cambridge University Press. <https://doi.org/10.1017/CBO9780511802157>.
- Giamellaro, M., Lan, M.-C., Ruiz-Primo, M. A., Li, M., & Tasker, T. (2017). Curriculum mapping as a strategy for supporting teachers in the articulation of learning goals. *Journal of Science Teacher Education*, 28(4), 347–366. <https://doi.org/10.1080/1046560X.2017.1343603>.
- Gipps, C. (1999). Chapter 10: Socio-cultural aspects of assessment. *Review of Research in Education*, 24(1), 355–392. <https://doi.org/10.3102/0091732x024001355>.
- Greeno, J. G. (2006). Learning in activity. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences*. Cambridge University Press.
- Gutiérrez, K. D., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, 32(5), 19–25. <https://doi.org/10.3102/0013189X032005019>.
- Haertel, E. H., Moss, P. A., Pullin, D. C., & Gee, J. P. (2008). Introduction. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 1–16). Cambridge University Press. <https://doi.org/10.1017/CBO9780511802157>.
- Hammond, Z. (2014). *Culturally responsive teaching and the brain: Promoting authentic engagement and rigor among culturally and linguistically diverse students*. Corwin: A Sage Company.
- Hammond, Z. (2021). Liberatory education: Integrating the science of learning and culturally responsive practice. *American Educator*, Summer. <https://www.aft.org/ae/summer2021/hammond>.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>.
- Herman, J. (2016). *Comprehensive standards-based assessment systems supporting learning*. National Center for Research on Evaluation, Standards, and Student Testing. [https://compcenternetwork.org/sites/default/files/archive/CAS\\_SupportingLearning.pdf](https://compcenternetwork.org/sites/default/files/archive/CAS_SupportingLearning.pdf).
- Jackson, P. W. (1971). The student's world. In M. L. Silberman (Ed.), *The experience of schooling* (pp. 9–25). Holt, Rinehart, and Winston, Inc.
- Jordan, B., & Putz, P. (2003). Thinking assessment. *Practicing Anthropology*, 25(4), 37–41. <https://www.jstor-org.stanford.idm.oclc.org/stable/24781425>.
- Kahlenberg, R. D., Potter, H., & Quick, K. (2019). School integration: How it can promote social cohesion and combat racism. *American Educator*, Fall 2019. [https://www.aft.org/ae/fall2019/kahlenberg\\_potter\\_quick](https://www.aft.org/ae/fall2019/kahlenberg_potter_quick).
- Kang, H., & Furtak, E. M. (2021). Learning theory, classroom assessment, and equity. *Educational Measurement: Issues and Practice*, 40(3), 73–82. <https://doi.org/10.1111/emip.12423>.
- Kang, H., Thompson, J., & Windschitl, M. (2014). Creating opportunities for students to show what they know: The role of scaffolding in assessment tasks. *Science Education*, 98(4), 674–704. <https://doi.org/10.1002/sce.21123>.
- Kang, H., Talafian, H., & Tschida, P. (2022). Expanding opportunities to learn in secondary science classrooms using unconventional forms of classroom assessments. *Journal of Research in Science Teaching*, 60(5), 1053–1091. <https://doi.org/10.1002/tea.21824>.
- Khisty, L. L., & Chval, K. B. (2002). Pedagogic discourse and equity in mathematics: When teachers' talk matters. *Mathematics Education Research Journal*, 14(3), 154–168. <https://doi.org/10.1007/BF03217360>.
- Killen, M., Elenbaas, L., & Rutland, A. (2015). Balancing the fair treatment of others while preserving group identity and autonomy. *Human Development*, 58(4–5), 253–272. <https://doi.org/10.1159/000444151>.
- Kisker, E. E., Lipka, J., Adams, B. L., Rickard, A., Andrew-Ihrke, D., Yanez, E. E., & Millard, A. (2012). The potential of a culturally based supplemental mathematics curriculum to improve the mathematics performance of Alaska Native and other students. *Journal of Research in Mathematics Education*, 43(1), 75–113.
- Kroog, H., King Hess, K., & Ruiz-Primo, M. A. (2016). The 2 Es: Implement effective and efficient approaches to formal formative assessment that will save time and boost student learning. Special Issue: Looking at Students' Work. *Educational Leadership*, 73(7), 22–25. <https://www.ascd.org/el/articles/the-2-es>.
- Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal*, 32(3), 465–491. <https://doi.org/10.3102/00028312032003465>.



- Ladson-Billings, G. (2014). Culturally relevant pedagogy 2.0: a.k.a. the remix. *Harvard Educational Review*, 84(1), 74–84. <https://doi.org/10.17763/haer.84.1.p2rj131485484751>.
- Ladson-Billings, G. (2021). *Culturally relevant pedagogy: Asking a different question*. Teachers College Press.
- Lampert, M., Franke, M. L., Kazemi, E., Ghouseini, H., Turrour, A. C., Beasley, H., Cunard, A., & Crowe, K. (2013). Keeping it complex: Using rehearsals to support novice teacher learning of ambitious teaching. *Journal of Teacher Education*, 64(3), 226–243. <https://doi.org/10.1177/0022487112473837>.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511815355>.
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute and day by day. *Educational Leadership*, 63(3), 18-24. <https://www.ascd.org/el/articles/classroom-assessment-minute-by-minute-day-by-day>.
- Learning in Places Collaborative. (2023). *For families*. <http://learninginplaces.org/for-families/family-le-1>.
- Learning in Places Collaborative. (n.d.). *Classroom storyline*. <http://learninginplaces.org/seasonal-storyline/classroom-storyline>.
- Lee, C. D. (1995). A culturally based cognitive apprenticeship: Teaching African American high school students skills in literary interpretation. *Reading Research Quarterly*, 30(4), 608–630. <https://doi.org/10.2307/748192>.
- Lee, C. D. (2008). Cultural modeling as opportunity to learn: Making problem solving explicit in culturally robust classrooms and implications for assessment. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 136–169). Cambridge University Press.
- Lipka, J. (2003). *Coming to egg island: Adventures in grouping and place values*. Detselig Enterprise Ltd.
- Lipka, J., Hogan, M. P., Webster, J. P. Yanez, E., Adams, B., Clark, S., & Lacy, D. (2005). Math in a cultural context: Two case studies of a successful culturally based math project. *Anthropology and Education Quarterly*, 36(4), 367–385. <https://doi.org/10.1525/aeq.2005.36.4.367>.
- McDermott, R., & Pea, R. (2020). Learning “how to mean”: Embodiment in cultural practices. In N. S. Nasir, C. D. Lee, R. Pea, & M. McKinney de Royston (Eds.), *Handbook of the cultural foundations of learning* (pp. 99–118). Routledge.
- McMillan, J. H. (2018). *Using students’ assessment mistakes and learning deficits to enhance motivation and learning*. Routledge: Taylor & Francis Group.
- Mehan, H. (2008). A sociological perspective on opportunity to learn and assessment. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 42–75). Cambridge University Press.
- 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](https://learningpolicyinstitute.org/media/111/download?inline&file=Social_Emotional_Learning_New_Accountability_REPORT.pdf).
- Michaels, S., O’Connor, M. C., Williams Hall, M., & Resnick, L. B. (2016). *Accountable talk® sourcebook: For classroom conversation that works*. Institute for Learning: University of Pittsburgh. <https://nsiexchange.org/wp-content/uploads/2019/02/AT-SOURCEBOOK2016-1-23-19.pdf>.
- Minstrell, J., Anderson, R., & Li, M. (2009). *Assessing teacher competency in formative assessment*. Annual Report to the National Science Foundation.
- Mohatt, G., & Erickson, F. (1981). Cultural differences in teaching styles in an Odawa school: A sociolinguist approach. In H. T. Trueba, G. P. Guthrie, & K. H-P. Au (Eds.), *Culture and the bilingual classroom: Studies in classroom ethnography* (pp. 108–119). Newbury House Publishers, Inc.
- Moll, L. C., & Greenberg, J. B. (1990). Creating zones of possibilities: Combining social contexts for instruction. In L. C. Moll (Ed.), *Vygotsky and education: Instructional implications and applications of sociohistorical psychology* (pp. 319–348). Cambridge University Press.
- Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory Into Practice*, 31(2), 132–141.
- Moss, C. M., & Brookhart, S. M. (2012). *Learning targets: Helping students aim for understanding in today’s lesson*. Association for Supervision and Curriculum Development.
- Moss, P. A. (2008). Sociocultural implications for assessment I: Classroom assessment. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 222–258). Cambridge University Press. <https://doi.org/10.1017/CBO9780511802157>.

- Murphy, P. K., & Alexander, P. A. (2007). Contextualizing learner-centered principles for teachers and teaching. In W. D. Hawley (Ed.), *The keys to effective schools: Educational reform as continuous improvement* (pp. 13–31). Corwin Press: A Sage Publications Company. <https://doi.org/10.4135/9781483329512>.
- Nasir, N. S. (2007). Identity, goals, and learning: The case of basketball mathematics. In N. S. Nasir & P. Cobb (Eds.), *Improving access to mathematics: Diversity and equity in the classroom* (pp. 132–145). Teachers College Press.
- Nasir, N. S., McKinney de Royston, M., Barron, B., Bell, P., Pea, R., Stevens, R., & Goldman, S. (2020). Learning pathways: How learning is culturally organized. In N. S. Nasir, C. D. Lee, R. Pea, & M. McKinney de Royston (Eds.), *Handbook of the cultural foundations of learning* (pp. 195–211). Routledge.
- Nasir, N. S., Rosebery, A. S., Warren, B., & Lee, C. D. (2022). Learning as a cultural process. Achieving equity through diversity. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (3rd ed., pp. 581–601). Cambridge University Press.
- National Academies of Sciences, Engineering, and Medicine. (2018a). *English learners in STEM subjects: Transforming classrooms, schools, and lives*. The National Academies Press. <https://doi.org/10.17226/25182>.
- National Academies of Sciences, Engineering, and Medicine. (2018b). *How people learn II: Learners, contexts, and cultures*. The National Academies Press. <https://doi.org/10.17226/24783>.
- National Academies of Sciences, Engineering, and Medicine. (2019). *Science and engineering for grades 6–12: Investigation and design at the center*. The National Academies Press. <https://doi.org/10.17226/25216>.
- National Academies of Sciences, Engineering, and Medicine. (2022). *Science and engineering in preschool through elementary grades: The brilliance of children and the strengths of educators*. The National Academies Press. <https://doi.org/10.17226/26215>.
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academy Press. <https://doi.org/10.17226/9853>.
- National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. National Academy Press. <https://doi.org/10.17226/10019>.
- Newman, D., Griffin, P., & Cole, M. (1984). Social constraints in laboratory and classroom tasks. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 172–193). Harvard University Press.
- Odden, T. O. B., & Russ, R. S. (2018). Sensemaking epistemic game: A model of student sensemaking processes in introductory physics. *Physical Review Physics Education Research*, 14(2). <https://doi.org/10.1103/PhysRevPhysEducRes.14.020122>.
- OpenSciEd. (2022). *OpenSciEd Teacher Handbook, Version 3.0*. OpenSciEd.org.
- Palmer, D. J., & Goetz, E. T. (1988). Selection and use of study strategies: The role of the studier’s beliefs about self and strategies. In C. E. Weinstein, E. T. Goetz, & P. A. Alexander (Eds.), *Learning and study strategies: Issues in assessment, instruction, and evaluation*. Academic Press.
- Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, 22(6), 806–813. <https://doi.org/10.1016/j.lindif.2012.04.007>.
- Panadero, E., Jonsson, A., & Strijbos, J-W. (2016). Scaffolding self-regulated learning through self-assessment and peer assessment: Guidelines for classroom implementation. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 311–326). Springer.
- Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher*, 41(3), 93–97. <https://doi.org/10.3102/0013189X12441244>.
- Penuel, W. R., & Shepard, L. A. (2016). Assessment and teaching. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (5th ed., pp. 787–850). American Educational Research Association.
- Penuel, W. R., & Shepard, L. A. (2017). Social models of learning and assessment. In A. A. Rupp & J. P. Leighton (Eds.), *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 146–173). John Wiley & Sons, Inc.
- Penuel, W. R., Fishman, B. J., Haugan Cheng, B., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher*, 40(7), 331–337. <https://doi.org/10.3102/0013189X11421826>.
- Penuel, W. R., Turner, M. L., Jacobs, J. K., Van Horne, K., & Sumner, T. (2019). Developing tasks to assess phenomenon-based science learning: Challenges and lessons learned from building proximal transfer tasks. *Science Education*, 103(6), 1367–1395. <https://doi.org/10.1002/sce.21544>.

- Pugh, S. L., Pawan, F., & Antommarchi, C. (2000). Academic literacy and the new college learner. In R. F. Flippo & D. C. Caverly (Eds.), *Handbook of college reading and study strategy research* (pp. 25–42). Lawrence Erlbaum.
- Randall, J. (2021). “Color-neutral” is not a thing: Redefining construct definition and representation through a justice-oriented critical antiracist lens. *Educational Measurement: Issues and Practice*, 40(4), 82–90. <https://doi.org/10.1111/emip.12429>.
- Randall, J., Poe, M., & Slomp, D. (2021). *Ain't oughta be in the dictionary: Getting to justice by dismantling anti-Black literacy assessment practices*. *Journal of Adolescent & Adult Literacy*, 64(5), 594–599. <https://doi.org/10.1002/jaal.1142>.
- Reiser, B. J., Novak, M., McGill, T. A. W., & Penuel, W. R. (2021). Storyline units: An instructional model to support coherence from the students' perspective. *Journal of Science Teacher Education*, 32(7), 805–829. <https://doi.org/10.1080/1046560X.2021.1884784>.
- Remillard, J. T., & Kim, O-K. (2020). *Elementary mathematics curriculum materials: Designs for student learning and teacher enactment*. Springer.
- Rogoff, B. (1995). Observing sociocultural activity on three planes: Participatory appropriation, guided participation, and apprenticeship. In J. V. Wertsch, P. del Rio, & A. Alvarez (Eds.), *Sociocultural studies of mind* (pp. 139–164). Cambridge University Press. <https://doi.org/10.1017/CBO9781139174299>.
- Rogoff, B. (1998). Cognition as a collaborative process. In W. Damon (Series Ed.) and D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 679–744). John Wiley & Sons, Inc. chrome-extension://efaidnbmnribpcajpcglclefindmkaj/<https://people.ucsc.edu/~brogoff/Scanned-articles/scanned%2012-2008/Cognition%20as%20a%20Collaborative%20Process.pdf>.
- Rogoff, B. (2003). *The cultural nature of human development*. Oxford University Press.
- Rogoff, B., & Angelillo, C. (2002). Investigating the coordinated functioning of multifaceted cultural practices in human development. *Human Development*, 45(4), 211–225. <https://doi.org/10.1159/000064981>.
- Rogoff, B., Paradise, R., Mejía Arauz, R., Correa-Chávez, M., & Angelillo, C. (2003). Firsthand learning through intent participation. *Annual Review of Psychology*, 54, 175–203. <https://doi.org/10.1146/annurev.psych.54.101601.145118>.
- Rosado-May, F. J., Urrieta, Jr., L., Dayton, A., & Rogoff, B. (2020). Innovation as a key feature of Indigenous ways of learning: Individuals and communities generating knowledge. In N. S. Nasir, C. D. Lee, R. Pea, & M. McKinney de Royston (Eds.), *Handbook of the cultural foundations of learning* (pp. 79–96). Routledge.
- Rosebery, A. S., Warren, B., & Tucker-Raymond, E. (2016). Developing interpretive power in science teaching. *Journal of Research in Science Teaching*, 53(10), 1571–1600. <https://doi.org/10.1002/tea.21267>.
- Rosebery, A. S., Solomon, F. C., Tucker-Raymond, E., Warren, B., Wright, C. G., Baumann, D., Godgrey, L., Solomon, J., & Fessenden, H. (n.d.). *Practice-based inquiry in science: A professional development course in science for teachers in K–5 districts*. TERC. <https://www.terc.edu/publications/practice-based-inquiry-in-science-a-professional-development-course-in-science-for-k-5-teachers-in-urban-districts>.
- Rosenholtz, S. J., & Simpson, C. (1984). The formation of ability conceptions: Developmental trend or social construction? *Review of Educational Research*, 54(1), 31–63. <https://doi.org/10.2307/1170397>.
- Ruiz-Primo, M. A. (2010). *Developing and Evaluating Measures of Formative Assessment Practice (DEMFAP) theoretical and methodological approach*. Internal manuscript. University of Colorado Denver. Laboratory of Educational Assessment, Research, and InnovationN.
- Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students' learning. *Studies in Educational Evaluation*, 37(1), 15–24. <https://doi.org/10.1016/j.stueduc.2011.04.003>.
- Ruiz-Primo, M. A. (2016). Implementing high-quality assessment for learning: Mapping as a professional development tool for understanding the what to learn, why to learn it, and how to learn it. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 219–236). Springer.
- Ruiz-Primo, M. A., (2021). *EDUC 364: Cognition and learning*. <https://explorecourses.stanford.edu/search?view=catalog&filter-coursestatus=Active=on&page=0&catalog=&q=EDUC364>.
- Ruiz-Primo, M. A., & Brookhart, S. M. (2018). *Using feedback to improve learning*. *Student Assessment for Educators* series. Routledge: Taylor & Francis Group.

- Ruiz-Primo, M. A. & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment, 11*(3–4), 237–263. <https://doi.org/10.1080/10627197.2006.9652991>.
- Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching, 44*(1), 57–84. <https://doi.org/10.1002/tea.20163>.
- Ruiz-Primo, M. A., & Kroog, H. (2018). Looking closely at mathematics and science classroom feedback practices: Examining artifacts, students' products, and teachers' communications. In A. A. Lipnevich & J. F. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 191–218). Cambridge University Press. <https://doi.org/10.1017/9781316832134>.
- Ruiz-Primo, M. A., Li, M., Birkby, E., Edwards, A., Occhipinti, K., Aaron, M., & Bair, M. (2013, April 6–9). *Looking at quality of instruction and students' performance: Where do the teachers' questions come from?* [Paper presentation]. National Association of Research in Science Teaching 2013 Annual International Conference, San Juan, Puerto Rico.
- Ruiz-Primo, M. A., Iverson, H., & Sands, D. (2014a, April 3–7). *Looking at feedback practices in science and mathematics classrooms: Lessons learned* [Paper presentation]. American Educational Research Association Annual Meeting, Philadelphia, PA, United States.
- Ruiz-Primo, M. A., Solano-Flores, G., & Li, M. (2014b). Formative assessment as a process of interaction through language: A framework for the inclusion of English language learners. In C. Wyatt-Smith, V. Klenowski, & P. Colbert, (Eds.), *Designing assessment for quality learning* (pp. 265–282). Springer.
- Ruiz-Primo, M. A., Kroog, H., & Sands, D. (2015, April 16–20). *Teacher-student interaction patterns: Looking at teachers with different levels of formative assessment expertise* [Paper presentation]. American Educational Research Association 2015 Annual Meeting, Chicago, Illinois.
- Ruiz-Primo, M. A., Sands, D., & Kroog, H. (2016). *Developing and evaluating measures of formative assessment practices (DEMFAP): Technical report on instrument development and evaluation*. Submitted to Institute of Education Sciences.
- Ruiz-Primo, M. A., Sands, D., & Kroog, H. (2022, April 21–26). *Microcultures: The role of classroom cultures in effective formative assessment* [Paper presentation]. American Educational Research Association 2022 Annual Meeting, San Diego, California.
- Ruiz-Primo, M. A., Kanopka, K., & Hernandez, P. (2024). Measuring, assessing, or just testing? Evidence from a few state testing programs in the U.S. In L. Cohen-Vogel, J. Scott, & P. Youngs (Eds.). *Handbook of education policy research* (2nd ed.). American Educational Research Association.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144. <https://doi.org/10.1007/BF00117714>.
- Schmidt, W. H., McKnight, C., Houang, R. T., Wang, H., Wiley, D. E., Cogan, L. S., & Wolfe, R. G. (2001). *Why schools matter: A cross-national comparison of curriculum and learning: The Jossey-Bass education series*. Jossey-Bass.
- Seidel, T., Rimmel, R., & Prenzel, M. (2005). Clarity and coherence of lesson goals as a scaffold for student learning. *Learning and Instruction, 15*(6), 539–556. <https://doi.org/10.1016/j.learninstruc.2005.08.004>.
- Shepard, L. A. (2019). Classroom assessment to support teaching and learning. *The ANNALS of the American Academy of Political and Social Science, 683*(1), 183–200. <https://doi.org/10.1177/0002716219843818>.
- Shepard, L. A. (2021). Ambitious teaching and equitable assessment: A vision for prioritizing learning, not testing. *American Educator*, Fall. <https://www.aft.org/ae/fall2021/shepard>.
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018a). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practices, 37*(1), 21–34. <https://doi.org/10.1111/emip.12189>.
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018b). Classroom assessment principles to support learning and avoid the harms of testing. *Educational Measurement: Issues and Practices, 37*(1), 52–57. <https://doi.org/10.1111/emip.12195>.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–22.
- Smith, J. B., Lee, V. E., & Newmann, F. M. (2001). *Achievement in Chicago elementary schools*. Consortium on Chicago School Research. <https://consortium.uchicago.edu/sites/default/files/2018-10/p0f01.pdf>.



- Smylie, M. A., & Wenzel, S. A. (2006). *Promoting instructional improvement: A strategic human resource management perspective*. Consortium on Chicago School Research. <https://consortium.uchicago.edu/sites/default/files/2018-10/p84.pdf>.
- Solano-Flores, G. (2016). *Assessing English language learners: Theory and practice*. Routledge: Taylor & Francis Group.
- Solano-Flores, G, Ruiz-Primo, M. A., Li, M., Zhao, X., Shade, C., & Chrzanowski, A. (2024). How equally do teachers distribute their attention across students classified as English learners (ELs) and their non-EL peer in science classrooms? A frequency analysis of monolingual and bilingual teachers' interactions with different student grouping configurations. *International Multilingual Research Journal*. <https://doi.org/10.1080/19313152.2024.2303275>.
- Spencer, M. B. (1999). Social and cultural influences on school adjustment: The application of an identity-focused cultural ecological perspective. *Educational Psychologist*, 34(1), 43–57. [https://doi.org/10.1207/s15326985ep3401\\_4](https://doi.org/10.1207/s15326985ep3401_4).
- Spindler, G., & Spindler, L. (1982). Roger Harker and Schönhausen: From familiar to stranger and back again. In G. Spindler (Ed.), *Doing the ethnography of schooling: Educational anthropology in action* (pp. 20–46). Waveland Press, Inc.
- Steel, D. M. (2012). Identity-safe school environments, creating. In J. A. Banks (Ed.), *Encyclopedia of diversity in education* (pp. 1125–1128). Sage.
- Stobart, G. (2005). Fairness in multicultural assessment systems. *Assessment in Education: Principles, Policy & Practice*, 12(3), 275–287. <https://doi.org/10.1080/09695940500337249>.
- Suárez, E. (2020). “Estoy Explorando Science”: Emergent bilingual students problematizing electrical phenomena through translanguaging. *Science Education*, 104(5), 791–826. <https://doi.org/10.1002/sce.21588>.
- Taylor, C. S., & Nolen, S. B. (2008). *Classroom assessment: Supporting teaching and learning in real classrooms* (2nd ed.). Pearson.
- Taylor, C. S., & Nolen, S. B. (2022). *Culturally and socially responsible assessment: Theory, research, and practice*. Teachers College Press.
- Taylor, M. C. (1979). Race, sex, and the expression of self-fulfilling prophecies in a laboratory teaching situation. *Journal of Personality and Social Psychology*, 37(6), 897–912. <https://doi.org/10.1037/0022-3514.37.6.897>.
- Tekumru-Kisa, M., & Stein, M. K. (2015). Learning to see teaching in new ways: A foundation for maintaining cognitive demand. *American Educational Research Journal*, 52(1), 105–136. <https://doi.org/10.3102/0002831214549452>.
- Tekumru-Kisa, M., Stein, M. K., & Schunn, C. (2015). A framework for analyzing cognitive demand and content-practices integration: Task analysis guide in science. *Journal of Research in Science Teaching*, 52(5), 659–685. <https://doi.org/10.1002/tea.21208>.
- Tekumru-Kisa, M., Stein, M. K., & Doyle, W. (2020). Theory and research on tasks revisited: Task as a context for students' thinking in the era of ambitious reforms in mathematics and science. *Educational Researcher*, 49(8), 606–617. <https://doi.org/10.3102/0013189X20932480>.
- Thompson, J. J., Hagenah, S., McDonald, S., & Barchenger, C. (2019). Toward a practice-based theory for how professional learning communities engage in the improvement of tools and practices for scientific modeling. *Science Education*, 103(6), 1423–1455. <https://doi.org/10.1002/sce.21547>.
- Tobias, S., & Everson, H. T. (2009). The importance of knowing what you know: A knowledge monitoring framework for studying metacognition in education. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 107–127). Routledge.
- Tzou, C., Meixi, Suárez, E., Bell, P., LaBonte, D., Starks, E., & Bang, M. (2019). Storywork in STEM-Art: Making, materiality and robotics within everyday acts of Indigenous presence and resurgence. *Cognition and Instruction*, 37(3), 306–326. <https://doi.org/10.1080/07370008.2019.1624547>.
- Tzou, C., Bang, M., & Bricker, L. (2021). Commentary: Designing science instructional materials that contribute to more just, equitable, and culturally thriving learning and teaching in science education. *Journal of Science Teacher Education*, 32(7), 858–864. <https://doi.org/10.1080/1046560X.2021.1964786>.
- Wang, T., Zhao, D., Birkby, E., Li, M., & Ruiz-Primo, M. A. (2013, April 26–30). *Knowledge of learning goals as a navigation tool in curriculum implementation* [Paper presentation]. 2013 National Council of Measurement in Education Annual Meeting, San Francisco, California.



- Warren, B., & Rosebery, A. S. (2011). Navigating interculturality: African American male students and the science classroom. *Journal of African American Males in Education*, 2(1), 98–115.
- Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 315–327). Macmillan Publishing Company.
- Wiliam, D. (2018). Feedback: At the heart of—but definitely not all of—formative assessment. In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 3–28). Cambridge University Press.
- Wiliam, D., & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Routledge.
- Windschitl, M., Thompson, J., & Braaten, M. (2018). *Ambitious science teaching*. Harvard Education Press.
- Winne, P. H. (1995). Inherent details in self-regulated learning. *Educational Psychologist*, 30(4), 173–187. [https://doi.org/10.1207/s15326985ep3004\\_2](https://doi.org/10.1207/s15326985ep3004_2).
- Winne, P. H., & Azevedo, R. (2014). Metacognition. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 63–87). Cambridge University Press. <https://doi.org/10.1017/CBO9781139519526>.
- Wright, C. G., Likely, R., Wendell, K. B., Paugh, P. P., & Smith, E. (2019). Recognition and positional identity in an elementary professional learning community: A case study. *Journal of Pre-College Engineering Education Research*, 10(1). <https://doi.org/10.7771/2157-9288.1214>.
- Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299–315). Routledge.

# Assessment Literacy and Professional Learning

E. Caroline Wylie, *National Center for the Improvement of Educational Assessment*

Margaret Heritage, *Heritage Consulting*

## CONTENTS

INTRODUCTION .....	133
ASSESSMENT LITERACY.....	134
Classroom Assessment.....	134
Assessment Literacy Knowledge and Skills: Learning Goals.....	134
Assessment Literacy Knowledge and Skills: Assessment Evidence .....	137
Assessment Literacy Knowledge and Skills: Interpretation of Student Responses and Action .....	140
ENABLING CONDITIONS FOR PROFESSIONAL LEARNING .....	144
Sociocultural Consciousness and Agency.....	144
Learning Supports .....	146
Deliberate Practice.....	147
PROFESSIONAL LEARNING ACTIVITIES FOR DEVELOPING ASSESSMENT LITERACY .....	148
Learning Goals.....	148
Elicitation of Assessment Evidence.....	151
Interpretation of Evidence and Action .....	154
SYSTEM SUPPORT FOR ASSESSMENT LITERACY .....	158
School and District Leaders .....	158
State Leaders .....	160
KEY TAKEAWAYS .....	161
REFERENCES.....	162

## INTRODUCTION

To be consistent with advances in the learning sciences, assessment must be reconceptualized. In Chapter 3 of this volume, “Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems,” Goldman and Lee observe that assessment should “reflect cultural, social, emotional, and cognitive dimensions” of human learning and development, in tandem with the traditionally prioritized cognitive dimension (p. 50). The purpose of this chapter is to consider what assessment literacy entails within this reconceptualization and how it can be promoted among teachers. Throughout this chapter, we position teachers as active agents in their own learning. Beginning in pre-service, candidate teachers should take an active stance and receive guidance from program faculty and cooperating teachers. Across their careers, they should become increasingly self-regulated, taking more control over their learning in collaboration with peers and school and district leaders (Heritage & Wylie, 2020).

In broad terms, we adopt the perspective that assessment literacy is the ability to engage in a chain of reasoning from evidence (Mislevy, 1994, 1996), a process that is always applicable regardless of the differing contexts, purposes, and timescales of assessment (Pellegrino, 2014). A chain of reasoning begins with identifying learning goals—what is to be assessed—followed by a means to elicit evidence of learning in relation to the goal and ends with interpreting evidence to guide asset-based and future-oriented actions to benefit student learning and development. This process characterizes all classroom assessment, from an end-of-unit assessment to an interaction between teacher and a student (Pellegrino et al., 2023).

A primary concern in addressing assessment literacy is how assessment can facilitate equitable and just learning outcomes for all students. We have adopted the definition of equity as “an approach to ensuring equally high outcomes for all by removing the predictability of success or failure that currently correlates with any racial, social, economic, or cultural factor” (Safir & Dugan, 2021, p. 29). Achieving equity requires a culturally sustaining approach to pedagogy and a fair and just approach to assessment, including interrogating the content of what is taught and how it is taught, together with what and how that content is assessed (Paris, 2012; Randall et al., 2022; Stemberge, 2020; Taylor & Nolen, 2022). Fair and just classroom assessment thrives to the extent that teachers and other professionals are also involved in culturally sustaining pedagogy; that they understand how to create assessment tasks that reflect students’ cultures, languages, and ways of knowing; and that they engage in equitable and just interpretations and actions based on information gained from the assessment.

This chapter is organized into three sections. First, we identify the knowledge and skills that teachers need to make effective use of classroom assessment. We then move from examining *what* teachers need to know to examining *how* they can develop competencies in assessment literacy. Next, we present a set of enabling conditions for teacher professional learning on assessment literacy, along with specific professional learning activities. The final section of this chapter addresses the role of school and district leaders and state policy in providing systemic support for assessment literacy.

## ASSESSMENT LITERACY

We begin this section by discussing classroom assessments that teachers can use to benefit their students' learning and development in the context of ambitious teaching. Then, we describe the knowledge and skills teachers need to make effective use of classroom assessment within an activity system. These competencies are organized around three components of reasoning from evidence: learning goals, eliciting evidence, and interpretation and action. To illustrate assessment literacy knowledge and skills in practice, we include an example of ambitious teaching and integrated formative assessment in a Grade 8 art lesson (see Tables 5-1, 5-2, and 5-3 later in this chapter).

### Classroom Assessment

Following Chapter 4 of this volume, "Classroom Activity Systems to Support Ambitious Teaching and Assessment," we locate classroom assessment within a learner-centered activity system that includes five integrated elements: learner, curriculum, instruction, assessment, and a classroom learning culture. These elements are grounded in ambitious teaching, which adopts a sociocultural approach to learning and centers on each learner's engagement and participation in rigorous learning opportunities. These opportunities connect to who students are and the knowledge and resources they bring to the classroom from their lived experiences in home and community (Shepard, 2021).

Within this activity system, classroom assessment is used for both formative and summative purposes, including grading. Formative assessment is "intimately connected with the process of teaching and learning" (Black, 1993, p. 51). Teachers gain insights into students' current learning status in order to guide ongoing teaching and learning decisions so that they can teach within the students' zone of proximal development (Torrance & Pryor, 1998). Students are prompted to develop metacognitive strategies so that they can purposefully direct their own learning.

Whereas formative assessment provides a steady stream of evidence to inform ongoing learning, classroom summative assessment gives a point-in-time view of achievement at the end of a period of learning—for example, the end of a unit or a course. Summative assessment results can be used to assign grades or otherwise certify achievement (Shepard, 2019), to inform future work, or to prompt further probing to understand weaker-than-expected performance among students.

To augment classroom-based assessment evidence, Safir and Dugan (2021) advocate for the use of "street data," information that comes from a variety of sources that include student interviews, identity maps, student ethnographies, home visits, and staff or student comment cards. Unlike test scores or other forms of summary data that provide a "satellite view" of achievement, street data provide an on-the-ground perspective "revealing students' assets, cultural wealth, and learning needs" (Safir & Dugan, 2021, p. 57) that can be used in conjunction with assessment evidence to provide real-time insights into the context surrounding student learning.

### Assessment Literacy Knowledge and Skills: Learning Goals

Learning goals—the foundation for both instruction and assessment—are rigorous, high-quality, meaningful, and challenging for students (Shepard, 2021). Clearly defined

success criteria help students understand what meeting a goal looks and sounds like. Learning goals and success criteria are developed from a combination of academic resources, teachers' disciplinary knowledge, and teachers' knowledge of their students. Academic resources can include learning standards, learning progressions that show a typical trajectory of learning, and learning sequences laid out in the curriculum. Teachers' disciplinary knowledge consists of the distinctive nature of the thinking processes and beliefs specific to a discipline; an understanding of how learning typically progresses in that discipline; and pedagogical content knowledge, "that special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding" (Shulman, 1987, p. 8). Using their knowledge of students, teachers need to tailor disciplinary learning goals to connect to their students' prior knowledge—from school, their cultural backgrounds, and their lived experiences (Moll et al., 1992; Sireci, 2020). Reflecting who students are in the classroom is important for engendering feelings of legitimacy, so that all students can feel safe and valued (Gutiérrez & Rogoff, 2003; Moll et al., 1992; National Academies of Sciences, Engineering, and Medicine, 2018). The specific knowledge and skills needed to develop learning goals and associated success criteria are presented in Box 5-1.

Table 5-1 focuses on setting up the Grade 8 art lesson, as well as communicating learning goals and success criteria. This lesson draws from an English language arts unit developed by Walqui et al. (2023) and observed by Heritage. The lesson was not recorded so student quotes reflect what was heard but are not taken from a transcript.

**BOX 5-1**  
**Specific Knowledge and Skills Needed to Develop**  
**Learning Goals and Associated Success Criteria**

- Knowledge of the distinct ways of knowing and reasoning that are specific to a discipline and how students come to learn in these distinct ways.
- Skills in creating goals that apprentice students to a discipline (e.g., behaving as a mathematician or a writer) while honoring and supporting their individual identities as learners.
- Knowledge of standards, curriculum sequences, and learning progressions within a discipline and/or skills to create learning progressions from the standards.
- Skills in creating worthwhile and rigorous learning goals that are aligned to standards, progressions, or curriculum materials that challenge each student based on their current learning status.
- Skills in describing success criteria to help students understand what success looks and sounds like.
- Knowledge of students' family and community beliefs, values, and culture, as well as the interests and gifts that students bring to the classroom.
- Skills in leveraging this knowledge to create learning goals that connect to students' lived experience and help students to gain insight into experiences different from their own.
- Knowledge of self-regulation, metacognition, motivation, and self-efficacy and their impacts on development.
- Skills in leveraging these constructs when making decisions about learning goals and success criteria (e.g., will student be motivated by this goal, or will all students be able to access this goal?).



The lefthand column describes how the lesson unfolded, while the righthand column draws attention to the specific aspects of assessment literacy that the teacher employed at that stage of the lesson.

Contrast the example in Table 5-1, showing a teacher modifying curriculum to better reflect her students’ culture and attend to their identity, with that of students who experience a curriculum that is agnostic to who they are and the resources they bring to the classroom. Assessment of student learning in this teacher-modified curriculum has the potential to “sustain, not eradicate, students’ cultures, languages, and ways of knowing/being” (Randall et al., 2022, p. 172), a hallmark of fair and justice-oriented assessment.

**TABLE 5-1** A Vignette Linked to Assessment Literacy Skills for Learning Goals

Classroom Practice	Teacher’s Assessment Literacy Knowledge and Skills
<p>This lesson comes from a Grade 8 class focusing on art as a form of storytelling—part of the visual literacy strand of the district’s art curriculum. The teacher planned a series of lessons based on these standards:</p> <p>Art: Develop Visual Literacy</p> <ul style="list-style-type: none"> <li>• Describe, analyze, and interpret created art</li> <li>• Speak and write effectively and clearly about works of art</li> </ul> <p>Many of the students’ families had roots in Mexico and the teacher knew that quite a few of them had also visited family there. To connect to some students’ Mexican heritage, and to broaden the cultural understanding of those students who did not share that heritage, the teacher began the lesson series with the Mexican artist Diego Rivera’s mural <i>Dream of a Sunday Afternoon in Alameda Park</i>. The mural depicts hundreds of characters from 400 years of Mexican history gathering for a walk in Mexico City’s largest park. The teacher wanted her students to see a powerful artist who shared their heritage and images of people who looked like them. The teacher would later help students apply what they had learned from analyzing this mural to other works of narrative art.</p>	<p>The teacher was able to set a goal, aligned to existing curriculum standards, that was challenging and meaningful to the students. She connected the goal to students’ cultural backgrounds because she knew they would be motivated by making connections to their families’ Mexican roots and would value learning about an artist that shared their heritage (based on the teacher’s knowledge of family background, student interest, and the neighborhood surrounding the school).</p>
<p>After introducing the focus of the lesson series, the teacher started the first lesson by asking students what they knew about murals and if they had seen murals anywhere in their community. Some of the students said they lived near a wall that had a lot of graffiti on it and thought that was a kind of mural. Some students said they had seen a wall painted with an image of Kobe Bryant. Other students said that on their way to school they passed Farmer John’s, a meat supplier, which had a mural of pigs in a field on the wall. The teacher asked the students what they thought the purpose of the various murals were, and their answers ranged from celebrating someone’s life to advertising what you were selling to showing that you were in a gang. From the discussion, students came to agree that the people who created a mural had a purpose and a message to communicate. After establishing this foundational knowledge, the teacher would now be able to draw on and make connections to the students’ prior knowledge about murals throughout the lesson.</p>	<p>The teacher capitalized on her knowledge of the students’ likely experiences with murals within their community.</p>

*continued*

**TABLE 5-1** Continued

Classroom Practice	Teacher’s Assessment Literacy Knowledge and Skills
<p>Then the teacher introduced the students to Diego Rivera, projecting his image on the whiteboard and giving some background about him. This background provided context for analyzing his mural—for example, how he favored mural painting because it could present subjects on a large scale to a wide public audience, consistent with his communist politics. The teacher then briefly introduced the class to key ideas of communism. She also noted that in 1922, after the Mexican Revolution, Rivera and others signed the <i>Manifiesto of the Syndicate of Technical Workers, Painters, and Sculptors</i>, arguing that artists must invest their greatest efforts to make art that was valuable to the people.</p>	<p>The teacher augmented the students’ background knowledge about Rivera so they could draw from it, as well as their local knowledge of murals, when they were analyzing <i>Dream of a Sunday Afternoon in Alameda Park</i>.</p>
<p>Next, the teacher asked the students to write individually in their journals about how they thought murals were different from other art forms they had studied. The teacher then led a class discussion in which students raised questions about art, including that they saw some forms of art being for rich people and only seen in museums, whereas ordinary people could view murals on the street. One student speculated that “maybe there is more of a story and a message in a mural like the ones we just talked about.” This prompted the teacher to ask, given the student’s background, what kinds of messages or stories the class thought Diego Rivera might have. Students offered ideas such as “stories about communism and poor people, messages from workers, stories from history.”</p>	<p>The teacher provided students with clear success criteria to help them understand what was expected of them, and the teacher made sure they understood both the goal and criteria before they began their task.</p>
<p>At this point, the teacher decided that the students had sufficient background knowledge to begin their analysis and projected a large image of Diego Rivera’s mural onto the whiteboard. The teacher explained that the goal of the lesson was to learn how to analyze an image—in this case, the Rivera mural. To reach that goal, the students would examine the details of the mural in sections and then discuss how they came together to tell a story. Their success criteria would be to describe what they saw in the mural, identify key information, and make inferences to explain what story the mural was telling. The teacher then asked the students to tell a partner in their own words what the goal and success criteria in the lesson entailed.</p>	

**Assessment Literacy Knowledge and Skills: Assessment Evidence**

Assessment evidence of learning for either formative or summative purposes comes from planned tasks or situations, aligned with learning goals that embody the cognitive and cultural dimensions of learning, and that prompt students to say, do, or create something that shows the status of their learning (National Research Council, 2001). The social and emotional dimensions of assessment are addressed by taking account of students’ motivations and interests, ensuring that students understand the purpose of the assessment, and confirming that students perceive the task as worthwhile and relevant (Shepard, 2000).

Assessment tasks or situations should offer multiple entry points and modalities in which knowledge and reasoning can be displayed—for example, tasks with differing levels of difficulty, oral and written language options, and both graphic and pictorial

representations (Nasir et al., 2021; Randall, 2021) so that all students can accurately show what they know. Assessment opportunities should also connect to and build on students' funds of knowledge, those knowledge assets that students have as a result of personal experiences in their homes, families, and communities (Esteban-Guitart & Moll, 2014; Moll et al., 1992; Subero et al., 2015). Students can also generate evidence of learning through their own internal feedback, or self-assessment—a process of comparing one's own performance to internally or externally provided criteria. In self-assessment, students form judgments about the level to which they have satisfied the criteria (Boud & Molloy, 2013) and make decisions about the actions they need to take next (Ames, 1992; Paris & Paris, 2001), including adapting learning strategies, revising work, or setting new goals. Of course, teachers need to support students in developing self-assessment skills through models and structures in the classroom. Helping students develop self-assessment skills does have “pay off” by increasing academic performance (Brown & Harris, 2013) and fostering metacognition, self-regulated learning, and self-efficacy (Panadero et al., 2016). The specific knowledge and skills teachers need for generating assessment evidence of learning are presented in Box 5-2.

**BOX 5-2**  
**Specific Knowledge and Skills Needed for Generating**  
**Assessment Evidence of Learning**

- Knowledge of the importance of coherence among learning opportunities, classroom formative and summative assessment (including grading), and how assessment purpose will inform how evidence of student understanding is produced.
- Skills in creating an optimal climate for learning and assessment, generating an atmosphere of trust and purpose, and ensuring collective orientation to learning and development.
- Knowledge and skills in planning situations, activities, tasks, or questions to elicit prior knowledge and evidence of progress toward the current learning goal(s) with shared indicators of successful performance that will be actionable in the here and now of learning, and, in the case of summative assessment, at the end of a period of learning.
- Knowledge of how to create assessment opportunities that sustain the specific local cultural and linguistic diversity present in each classroom and support students' ways of knowing and being.
- Skills in planning authentic and worthwhile tasks with multiple modes (e.g., written, oral, performance) that require students to engage with powerful disciplinary ideas and practices that incorporate their funds of knowledge they bring to school from their homes and communities; that have sufficiently broad entry points to provide all students with the opportunity to show where they are in their learning in ways that situate them as competent; and that are accessible to students with disabilities and who are English learners.
- Skills in ensuring metacognitive skill development (including goal setting and self-monitoring) and promoting the ongoing use of these skills in the classroom to help students understand their own learning status and performance.
- Skills in identifying and collecting other sources of information (i.e., street data) to support deeper insights into student learning compared to solely considering assessment data.

Table 5-2 continues the example of the Grade 8 art class with a focus on how the teacher elicits evidence of student learning throughout the lesson.

**TABLE 5-2** A Vignette Linked to Assessment Literacy Skills for Eliciting Assessment Evidence

Classroom Practice	Teacher’s Assessment Literacy Knowledge and Skills
<p>When the teacher thought that the students understood the learning goal and success criteria, she gave images of four sections of the mural to each group of four students to analyze. First they were to work individually on one section each and then share their thinking with each other in order to decide together what story Rivera was telling with the complete mural. To scaffold this analysis, the teacher provided questions for individual students to write responses to before they came together for a group discussion—for example, “What do you see in the image? What stands out to you? How are people in the mural interacting? What did Diego Rivera want to convey to people and why do you think that?” The students could respond in English or a combination of English and Spanish.</p>	<p>The open-ended questions that the teacher asked to capture students’ preliminary ideas about the mural in their notes, along with their more refined understandings as a result of the group discussions, provided the teacher with evidence of student understanding.</p>
<p>As the students were completing their individual writing tasks, the teacher observed the students’ work and, in instances where a student’s writing was limited, prompted them with questions like, “What are the details you notice in this section? Why do you think Rivera introduced these images? What do you think he was trying to say? What title would you give this section? What do you see that makes you say that?” As students responded to these prompts, the teacher obtained more evidence about their analytic thinking and the students had a chance to deepen their analysis.</p>	<p>The use of multiple modalities and translanguaging helped the teacher access the emerging thinking of her bilingual students while sustaining linguistic diversity in the classroom.</p>
<p>After their individual writing tasks were complete, the students placed each section of the mural side by side so the small groups were able to see the full image. Before they began their small group discussions, the teacher reminded them that they are learning to describe, analyze, and interpret art and encouraged them to be explicit about whether they were offering a description, analysis, or interpretation. The students shared their ideas about their respective sections in their groups using discussion prompts to help start the discussion, including, “Which objects stand out to you? What do you think they represent? What is a question you have about the mural or about Diego Rivera?” Some students added to their responses or revised their ideas based on what their peers said. While the students were engaged in their small group conversations, the teacher listened and asked probing questions to gain insights about their thinking, including, “What did your classmate say that made you change your idea? Why did that particular object stand out to you? What do you think Rivera intended by including it?”</p>	<p>The assessment evidence of student understanding was aligned with the learning goal and was proximal to the learning itself. The teacher used the learning goal throughout the discussion to help students connect the specific case of the Rivera mural to broader art appreciation skills.</p> <p>Because the teacher had cultivated a classroom climate in which students felt safe to express their ideas in English or Spanish, regardless of language proficiency, and where students recognized the value in listening to their peers’ ideas, the classroom was an optimal environment for the teacher to elicit evidence of current understanding from her students.</p>

### **Assessment Literacy Knowledge and Skills: Interpretation of Student Responses and Action**

The interpretation of student responses to tasks and situations in order to guide future action requires evidentiary reasoning based on disciplinary content knowledge (Bennett, 2019) and the teacher's knowledge of their students. Analysis of the evidence obtained from assessments should be used to develop an asset-based explanation or interpretation of the qualitative and quantitative assessment information and guide decisions on how best to advance every student's learning and development. Asset-based interpretations necessitate a shift from a reductive binary categorization of students—"got it" or "didn't get it"—to a more fine-grained view that identifies the specifics of what students understand and do not yet understand.

Interpreting student responses in a manner that integrates the cultural dimension of learning requires a focus on equitable and just interpretation of evidence stemming from teachers' sociocultural consciousness—for instance, teachers need to be careful to not privilege any students' linguistic and cultural patterns and practices that are more aligned with their own (Randall et al., 2022). Interpretation that considers the social and emotional dimension of learning is also informed by the teacher's knowledge of students' cultures, lived experiences, and current learning needs (Safir & Dugan, 2021); their knowledge of students' attitudes to and interest in the instructional and assessment task content (Ames, 1992); and their knowledge of the students' sense of self-efficacy with regard to the discipline (Bandura, 1977, 1993). An augmented picture of student learning performance based on a teacher's deep knowledge of their students and epistemological resources optimizes the potential for sensitive action that builds on students' current strengths and sustains their learning within the context of their language, literacies, and cultural ways of being (Paris & Alim, 2017).

Interpretation of assessment evidence must be followed by action by the teacher or student. Several carefully designed studies have demonstrated a positive impact on student learning when teachers use assessment evidence to make instructional adjustments (Bergan et al., 1991; Fuchs et al., 1991). Action based on interpretation from a cognitive perspective should be tailored to students' academic knowledge, skills, and analytic practices; and can take varied forms, including offering additional scaffolding to support deeper learning, sharing ideas and approaches from other students, introducing a new learning activity, using metaphor or representations, or providing feedback to guide revision and reflection with sufficient time to process and respond to that feedback. Longer-term adjustments, likely based on analysis of summative assessment evidence or data across multiple sources, may include modifying an upcoming unit to provide opportunities for some students to revisit a concept they have not yet fully grasped or examining trends across students or classrooms to inform grade- or department-level pedagogical, curricular, or assessment modifications.

It is critical that teachers use diverse student ideas and experiences, sourced from assessment evidence, as starting points for navigating between everyday forms of knowing and those forms of knowing that are accepted and used within specific content areas (Au & Kawakami, 1994; Bang & Medin, 2010; Cowie et al., 2018). This approach is especially important in formative assessment.

Action based on interpretation that addresses the cultural aspect of learning may include determining that students would benefit from a stronger integration of funds



of knowledge into future learning activities. For example, in a social studies unit, Ms. Cárdenas was teaching a Grade 2 class of English learners who were exploring their interest in civil rights, which had been piqued by a workers' strike occurring in their neighborhood. After the class's initial discussions and reading about rights—including *The Youngest Marcher* (Levinson & Brantley-Newton, 2017), about the youngest known child to be arrested at a civil rights protest in Birmingham, Alabama in 1963—the teacher wanted to strengthen the students' understanding by connecting rights to their own lives. She invited them to make clay models related to what they perceived as their personal rights and explain them. The students' explanations included: *I want to always have the right to live with my brother because in some places today families are separated. I have the right to be bilingual because if I'm not bilingual I cannot do things like talk to my grandma and read in more than one language* (personal communication, April 2017).

Action based on interpretation that addresses the social and emotional dimension of learning helps students regard assessment as a means for learning (Pryor, 2010) as they receive ongoing improvement-oriented feedback (Duijnhouwer et al., 2010). Such feedback offers students specific and actionable suggestions they can use—or not use, since feedback is not always a mandate. It focuses on the learning—the task—and not on the learner, which may lead students to set or revise their own goals, promoting feelings of competence (Andrade & Heritage, 2017). The specific knowledge and skills needed for the interpretation of student responses and action are presented in Box 5-3.

We conclude the example of practice from the Grade 8 art class in Table 5-3 by examining how the teacher interpreted evidence from students and acted on the insights she gained. While the separate tables help to explicate the different aspects of assessment literacy, the divisions are artificial. In other words, the teacher was collecting evidence of student understanding even during the initial stage of introducing the learning goal, which informed her decision to begin the main part of the lesson (see Table 5-1). Eliciting evidence, interpreting it, and taking subsequent action also all happened in close temporal proximity—for instance, when the teacher observed what students were writing and then asked them additional questions to help them deepen their observations of the mural (see Table 5-2). Interpretation and action based on assessment evidence most effectively supports learning when it occurs in the ongoing flow of a lesson—one or more class periods—and not as a distant event after learning has been completed (Black & Wiliam, 1998; Shepard, 2021).

The art teacher used ambitious teaching practices in this lesson, incorporating students' interests, backgrounds, and experience in an authentic inquiry. She built on students' prior knowledge and engaged them in learning as a social process, using appropriate scaffolds to support the entire class's participation so that they could develop visual literacy knowledge and skills, analytic abilities, and language skills by working collectively in a learning community.

The teacher's formative assessment practices were undergirded by her assessment literacy knowledge and skills. She designed multiple assessment opportunities into her teaching so that she could gauge how learning was developing across the class period and take contingent action. Representing understanding was not restricted to one mode—she supported her students' communicating their understanding in a variety of ways, including encouraging emergent bilingual students to use both Spanish and

**BOX 5-3**  
**Specific Knowledge and Skills Needed for**  
**Interpretation of Student Responses and Action**

- Skills in asset-focused evidentiary reasoning based in disciplinary content knowledge, recognizing strengths in student performance in order to determine next steps based on interpretation of formative assessment evidence, and evaluating student achievement based on interpretation of summative data.
- Knowledge of sources of corroborating, complementary, or other collections of data (e.g., street data) to provide a broader and deeper interpretation and understanding of learning and development, including knowledge of students' backgrounds, cultural frames of reference, interest and motivation in learning, and personal circumstances.
- Conscious knowledge of possible assumptions or biases and skills in minimizing them when making interpretations about student learning, recognition of whose voices are frequently marginalized, and skills to be more inclusive when collecting other sources of information to contribute to a well-rounded picture of students and their strengths and areas in which they need support.
- Skills in planning contingent responses based on student needs inferred from evidence, including leveraging student ideas as bridges to content area concepts; using improvement-oriented feedback generated by the teacher, peers, or the individual student's self assessment; and giving students time to use it by structuring additional activities for student discourse to advance learning.
- Knowledge of fair and effective grading practices and when they should be appropriately applied (i.e., not in the context of formative assessment).
- Skills in using interpretations from summative assessments to inform evaluation of curricular units, teaching practices, performance of subgroups of students, and trends across and among classrooms.
- Knowledge of the cultural components needed to advance learning, skills in weaving specific cultural aspects of students' backgrounds into teaching and learning, and skills in prompting students to draw on their funds of knowledge during sense-making.
- Skills in collaborating with students to understand learning status and performance in ways that enhance feelings of self-efficacy; and showing students how their responses shaped next steps.
- Knowledge of self-regulation and its impact on learning and motivation, and skills in teaching self-regulated learning processes.

English. All students were positioned as competent, with personal experiences that they could share, and each student's contribution was recognized in the paired work and the class discussion. The teacher used the evidence obtained from students' writing, discussions, and responses to take asset-based and future-oriented actions intended to move each student's learning forward.

In the example communicated through Tables 5-1, 5-2, and 5-3, the teacher's assessment was solely formative and, as a result, two key assessment literacy skills were not illustrated: grading and the relationship between classroom summative and formative assessment. Noting the problems with many grading practices, Chapter 4 of this volume, "Classroom Activity Systems to Support Ambitious Teaching and Assessment," emphasizes that grades should be based solely on what students know and can do,

**TABLE 5-3** A Vignette Linked to Assessment Literacy Skills for Interpreting and Taking Action on Evidence

Classroom Practice	Teacher’s Assessment Literacy Knowledge and Skills
<p>Based on her observations, the teacher concluded that while students could describe what they were seeing in the mural and what stood out for them, most groups could not yet explain what story they thought Rivera was telling.</p>	<p>The teacher focused on what students could do and continuously progressed their learning by taking action that matched their current learning status.</p>
<p>In response to where she thought the students’ thinking was, the teacher helped their analysis by prompting them to consider an emotion or feeling they had about the mural and the reason why. To scaffold their thinking, the teacher asked the students who initially mentioned the mural of Kobe Bryant to describe how they felt when they saw it. One student shared that he felt proud because he was a Lakers fan. Another student said it made her feel motivated to keep practicing with her basketball team. Another student said it made him feel sad about how Kobe had died. The teacher then invited the students to think about an emotion they had about the Rivera mural and share with a peer. After their paired conversation, the teacher led a class discussion where students volunteered their ideas and identified any details in the mural that contributed to their emotional reaction. For instance, some thought a smiling skeleton in the middle of the mural was scary and didn’t understand why it was there. Others talked about the violent incidents they observed in the mural involving Indigenous people and why those made them feel angry.</p>	<p>The teacher drew on earlier discussions about local murals to help students connect the emotions created by the Rivera mural to the story it was telling.</p>
<p>She then invited the students to share their ideas about the mural’s story. She guided the discussion so that students could build on each other’s ideas. As the discussion progressed, students added to each other’s perspectives and sometimes made alternate suggestions. For each idea presented, the teacher asked the student to refer to the mural for its source. The main ideas that surfaced were that Rivera wanted to show different people throughout Mexican history, both rich and poor, and that some poor people were not treated well. Some students inferred that Rivera thought that people would do better under communism, an idea that was picked up by other students in the classroom after they had heard their peers express it.</p>	<p>The teacher gave the students an opportunity for reflection on their learning and used evidence she obtained from both that reflection and the lesson itself to plan next instructional steps intended to deepen and expand their understanding of the Rivera mural, and then learn about other Mexican muralists.</p>
<p>At the conclusion of the lesson, the teacher returned to the learning goals and let students know that they would continue to develop their descriptive, analytical, and interpretative skills on other works of art in future lessons. She then asked each student to complete an exit ticket and respond to the questions, “What was your key takeaway from today’s lesson? What do you think you need help with? What would you like to learn more about?”</p>	
<p>Based on the final class discussion and review of the individual exit tickets, the teacher decided that as a next step she would invite the small groups to reconvene and come up with three questions that would help them better understand the message and story of the mural. Their questions eventually led the students to learn more about the Mexican Revolution, communism, and other Mexican muralists with similar views to Rivera.</p>	

and not on any other extraneous criteria. This point is underscored by Feldman (2019), when he stresses that equitable grading that is “accurate and bias-resistant includes *nothing other than a student’s summative assessment results*” (p. 143, italic in original). If the teacher in Tables 5-1, 5-2, and 5-3 was planning to assign grades, she would have done so at a later point based on summative assessment—for example, an end-of-unit assessment. Ideally, the summative assessment would be created from the macro goals of the unit, from which micro lesson goals for formative purposes were derived, so as to ensure synergy between the two forms of assessment. In addition to using the summative results to assign grades, the teacher would be able to use that information to make decisions about future unit content and any necessary pedagogical changes.

## ENABLING CONDITIONS FOR PROFESSIONAL LEARNING

In this section, we discuss three enabling conditions for professional learning—sociocultural consciousness and agency, learning supports, and deliberate practice—that ground teachers’ professional learning to develop assessment literacy competencies, regardless of the stage of their career (see Figure 5-1). In Figure 5-1, assessment competencies—the focus of professional learning across each of the three identified enabling conditions—are subsumed under the broad headings of learning goals, assessment evidence, and interpretation and action. Ambitious teaching is at the center of the figure since it is foundational for equity-focused assessment (Shepard, 2021) and is the context in which teachers make use of their assessment literacy knowledge and skills to benefit learning and development.

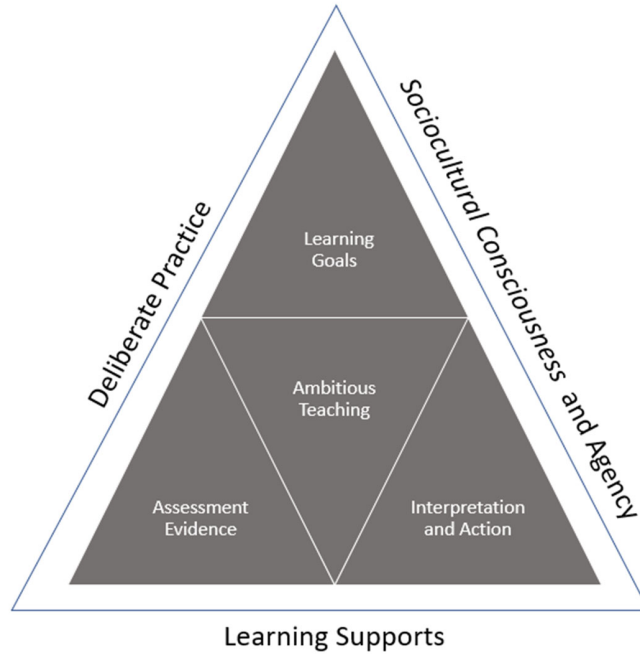
We expect that, for the most part, these enabling conditions will be operationalized in teachers’ local settings so that they can collaborate with their peers on continuous improvement of their assessment literacy knowledge and skills. In addition, each enabling condition should be supported by the direct involvement of school and district leaders, who play a pivotal role in helping teachers develop assessment literacy competencies (Stiggins & Duke, 2008).

### Sociocultural Consciousness and Agency

In this section, we describe how developing sociocultural consciousness and supporting teacher agency are critical for professional learning.

#### *Sociocultural Consciousness*

To engage meaningfully in equitable assessment, teachers must understand that their worldview is not universal, but has been profoundly shaped by their life experiences and mediated by a variety of factors—chief among them race, ethnicity, social class, and gender (Villegas & Lucas, 2001). Part of this process of understanding involves teachers recognizing the ways in which privilege and power operate in society in general, and within school systems in particular. Developing this understanding is the basis of sociocultural consciousness and requires teachers to critically reflect on their individual attitudes, beliefs, and values related to students and their backgrounds, schooling, and assessment (Heritage & Wylie, 2020). Such reflection sensitizes individuals to their



**FIGURE 5-1** Content and enabling conditions for developing assessment literacy.

own social identities and relationships to power, which bear on their work in schools and their local communities (Randall et al., 2022) and help to counter artificially low expectations of traditionally minoritized students, increase educators’ understanding of students’ lived experiences, and inform asset-based interpretations of assessment evidence of student learning.

Teachers may also work with colleagues to develop sociocultural consciousness—for instance, by discussing the cultural facets they have in common and how these may differ from those of their students. Such conversations can lead to considerations of how these differences might impact their attitudes and behaviors toward their students and how to ameliorate these attitudes and behaviors. Teachers may also want to read and discuss resources written by traditionally marginalized individuals and groups that provide perspectives on race, culture, and language that differ from their own.

It is equally important for school and district leaders to engage in comparable reflections so that they can lead explorations of assessment practices from an equitable and just perspective (Villegas & Lucas, 2001) and ensure that assessment practices are undergirded by equity-focused curricula, standards, and pedagogy across the school and district. In this vein, Marvin Pryor (personal communication, November 2022), former principal of The New Schools at Carver in Atlanta, Georgia, remarked that he led his school with the belief that “all students will learn under our care, not *can* learn, but *will* learn.”



### *Teacher Agency*

Teacher agency is enhanced when professional learning is treated as an inquiry process, in which teachers bring their problems of practice to a community of peers for exploration, reflection, and feedback. Such problems of practice could encompass the full range of assessment literacy competencies outlined earlier in this chapter. Although school and district leadership might identify areas of professional learning for which they have evidence of a system-wide need, the inquiry process encourages teachers to be active problem-solvers rather than only recipients of expert knowledge (Calvert, 2016). Furthermore, to preserve teachers' role as active agents, school and district leadership should ensure that professional learning activities help teachers achieve their personal goals and provide access to opportunities that are genuinely differentiated according to teachers' needs or expertise (Goe et al., 2017).

### **Learning Supports**

In this section, we describe two specific learning supports that can be used to develop each of the assessment competencies: access to a learning community and to expertise.

#### *Learning Community*

By providing a forum for participants to come together and deepen their knowledge and expertise by interacting on an ongoing basis (Wenger et al., 2002), learning communities are fundamental to nurturing teacher agency. In a learning community, participants take intentional and responsible management of their learning, utilize others as a resource for their own learning in the context of their own curricula, contribute to the growth of their peers, and act in new and creative ways (Calvert, 2016; Toom et al., 2015). Learning communities should exist for a sufficient enough duration that teachers have time to learn, practice, incorporate new ideas into their regular teaching practice, and reflect with colleagues on their implementation (Darling-Hammond et al., 2017).

Many schools have dedicated blocks of time for department- or grade-level teams to learn together—either short blocks of time on a regular basis or less frequent but longer blocks of time, such as a monthly early closure for students so that teachers can meet for the afternoon. Teachers should determine the assessment literacy focus for these blocks of time and combine learning and practical application into their conversation. School leaders play a role in ensuring that these blocks of time are preserved for this purpose.

An example of the value of a learning community comes from a case study of high school teachers focused on formative assessment. The participating teachers used no-carbon-required (NCR) paper, so that researchers would have a copy of their plans, to record how they were planning to try new strategies or continue with others they are familiar with after receiving feedback from their group, or address other aspects of formative assessment practice (Wylie et al., 2009). In an interview, one teacher described how the expectation of committing to try out something new in his classroom and then reporting on it created an informal, but powerful, sense of accountability:

BUT I'm sitting with my friends and on the NCR form I write down what I am going to do next month. Well, it turns out to be a sort of "I'm telling my friends I'm going to do this" and I really actually did it and it was because of that...by the next month you better take out that piece of paper and say "did I do that" and even if you didn't do it, you KNEW that you made a commitment to do [it] ... the idea of sitting in a group, working out something, and making a commitment, even something as informal [as writing on the NCR paper] I was impressed about how that actually made me do stuff. (Wylie et al., 2009, pp. 24–25)

### *Access to Expertise*

While school-based learning communities provide a valuable forum for learning, when a group is at a novice stage in their collective assessment literacy, members may struggle to accurately attend to the most important aspects of their own practice and peers may not yet know how best to press their colleagues to reflect critically (Sherin et al., 2011). In this situation, injecting sustained expertise into the learning community can be useful. For instance, coaches or teacher leaders can serve critical, expert roles such as instructional specialist, curriculum specialist, classroom supporter (teaching demonstration lessons, co-teaching, or observing to provide feedback), learning facilitator, or mentor to support teachers' assessment literacy in a range of contexts (Heritage & Wylie, 2020).

In addition to in-person expertise, other external supports that can assist learning communities include video or written examples of classroom practice, which are most instructive when accompanied by analysis that draws attention to critical features of the example. Disciplinary content expertise can be developed through the use of resources like research-based learning progressions and practical applications like deconstructing standards to examine the sub-goals while simultaneously considering how standards combine into major disciplinary ideas (Heritage & Wylie, 2020).

### **Deliberate Practice**

Deliberate practice, the third enabling condition for professional learning, entails specific and sustained efforts to do something that a person cannot do well (Ericsson et al., 1993). In summary, the characteristics of deliberate practice—in any area, not only teaching—are (1) a motivated individual who is attending to a task at hand and willing to exert effort to improve; (2) a scaffolded task that takes into account the prior learning of the subject; (3) the opportunity for brief instruction to support performance of the task; (4) the provision of informative feedback to the subject about their performance; and (5) repeatedly undertaking similar tasks over time (Ericsson et al., 1993). Classroom observation is a way to support deliberate practice, applicable to developing the range of assessment literacy knowledge and skills outlined previously.

### *Classroom Observation of Assessment Practices*

Classroom observation—either in-person or from a video recording—of assessment practices with feedback from a peer or a coach can complement work done in a learning community and permit teachers to exercise agency by directing the focus of the

observation (Wylie & Lyon, 2020). For instance, a teacher might ask a peer to attend to the quality of classroom questioning, noting to whom the teacher directs questions, the nature of each question, how students respond, and the pedagogical action the teacher takes. Discussion after the lesson can address how questioning informed teacher insights into student understanding as the lesson unfolded and ways in which future questioning can be improved (Wylie & Lyon, 2020). The observed teacher could practice implementing improvements and then request a subsequent observation to discuss the impact of their efforts, and so on. Tools can also support this form of deliberate practice—for example, an observation protocol with rubrics for various dimensions of formative assessment, including a template for improvement planning based on feedback discussions between peers (Wylie & Lyon, 2016).

Principals or other administrators can also promote deliberate practice by observing assessment practice in classrooms and having conversations with teachers to support reflection (Stronge & Xu, 2021). These observations and conversations could also serve as the basis for constructive feedback at a department or school level, if applicable to all teachers. Such cross-school or grade-level observations might focus on how students' epistemologies are used in instruction and assessment, the degree to which multiple modes of assessment are employed in a lesson or unit, and the use of evidence to advance learning. If a specific area of improvement emerges from these observations, teachers could engage in cycles of deliberate practice, implementing new approaches, reviewing them together, and making refinements.

## **PROFESSIONAL LEARNING ACTIVITIES FOR DEVELOPING ASSESSMENT LITERACY**

In this section, we describe specific professional learning activities for applying sociocultural consciousness and agency, learning supports, and deliberate practice to developing assessment literacy skills. These activities are not intended as a list or to be worked through exhaustively and in order. Rather, when teachers and those who support them decide on a particular assessment literacy focus area, they can draw from the suggestions below to match their identified needs.

### **Learning Goals**

#### *Developing and Refining Learning Goals*

Developing expertise in effectively creating, modifying, or utilizing learning goals for the purpose of assessment and instruction is a continuous process of review and refinement. A case in point is two experienced and skilled formative assessment practitioners who report that they still “share them [learning goals] with one another and get feedback because they're not always one hundred percent” (Heritage & Wylie, 2020, p. 209).

Collaboratively analyzing curriculum materials to identify the progression of concepts and analytic practices therein can support the development of learning goals and help deepen teachers' disciplinary knowledge. Similarly, learning goals can be improved by examining existing progressions—for example, the progression of science practices

in the Next Generation Science Standards Appendix E<sup>1</sup>—and creating local progressions derived from existing standards, tracing intermediate learning steps between standards in adjacent grade levels. For an example of how to create teacher-developed progressions, see Heritage (2021). Collaboratively developing or modifying lesson or unit goals with certain questions in mind, such as those in Box 5-4, can strengthen the skills of individual teachers in developing and refining learning goals.

**BOX 5-4**  
**Questions to Guide the Development of Learning Goals**

- Are the goals aligned to the standards, progression, or curricular materials?
- Are the goals rigorous for all students?
- Do the goals apprentice students to the discipline?
- Do the goals build on, and are they coherent with, students' prior academic learning?
- Do the goals combine cognitive, social and emotional, and cultural dimensions—for example, by reflecting students' family- and community-based funds of knowledge and nurturing students' identities?

Teachers can also examine whether the learning goals address cognitive, social and emotional, and cultural dimensions. Identifying some exemplar goals that address the social and emotional and cultural dimensions can then inform expansions and revisions to current curriculum learning goals. For example, a curricular goal for students was to “understand the way of life” depicted in Ernest Hemingway’s novella *The Old Man and the Sea*. Teachers modified this goal so that it read: “To understand the way of life depicted in the novella, compare what is important in your life to what is important in the life of Santiago [character in the novella].”<sup>4</sup>In addition to the cognitive dimension (understanding the way of life), the revised goal incorporated the emotional dimension (thinking about one’s own life in relation to someone else’s), and the cultural dimension (what is important in their own lives) (Heritage, 2021). Enabling students to work with a partner to share ideas about what is important in their lives compared to Santiago’s life adds the social dimension to the activity. A useful approach to ensuring these dimensions are embedded in learning goals might be to work on an upcoming unit, review how students responded to the goals, and then identify potential revisions for the following year, before moving on to another unit.

It is also useful for teachers to review and reflect on learning goals at the end of a lesson or unit to evaluate how well they worked, using questions such as those in Box 5-5. A personal reflection can sometimes be sufficient for such an analysis, but if multiple teachers have taught a lesson or unit with the same learning goals, they would likely benefit from a collaborative review.

---

<sup>1</sup> See <https://www.nextgenscience.org/sites/default/files/resource/files/AppendixE-ProgressionswithinNGSS-061617.pdf>.

**BOX 5-5**  
**Questions to Guide the Analysis of Learning Goals**

- Did the learning goals embody effective disciplinary representations (concepts and analytic practices)?
- Did the learning goals lead to rich, productive learning experiences?
- Were the learning goals accessible and meaningful for all students?
- Did the learning goals effectively build on students' prior learning, including their lived experiences?
- Were the learning goals the appropriate grain-size for a unit or a lesson?

*Establishing Shared Assessment Criteria*

Clarity of criteria is foundational for good assessment no matter whether it is summative, including grading, or formative. Teachers' skills in this area can be enhanced through deliberate practice for developing, trying out, and revising success criteria for learning goals—what the students will say or do to show they have reached the goal—following an inquiry process such as “Plan-Do-Study-Act” cycle (Russell et al., 2020). Teachers will need to draw on their disciplinary knowledge for this purpose and ongoing deliberate practice is necessary to develop expertise in establishing success criteria. Resources such as the Next Generation Science Standards Evidence Statements<sup>2</sup> can be a useful starting point to create lesson-level success criteria. Outside or within a learning community, teachers can reflect on success criteria after teaching a lesson or unit to determine how well the criteria worked for formative or summative assessment purposes and make further refinements if necessary.

*Incorporating Funds of Knowledge*

Before teachers can incorporate students' funds of knowledge into lesson or unit goals, they must have some familiarity with the beliefs, values, and practices of the communities to which their students belong. To acquire this knowledge, teachers can work together to create an ethnography of their school community, drawing on interviews with families and community leaders about the demographics of the area, the heritage of local families, common languages, religious observances, food, and local industries and businesses. Teachers can also understand students' interests and preferred activities at home and in the local community by asking them to produce identity artifacts—texts or drawings—which can help teachers access their funds of knowledge (Esteban-Guitart & Moll, 2014; Subero et al., 2015). Similarly, providing an exit ticket to students at the end of a lesson can probe the degree to which students found the lesson meaningful and relevant to local concerns, as well as their affective response to the lesson (Raza et al., under review).

---

<sup>2</sup> See <https://www.nextgenscience.org/evidence-statements>.



Reading about other teachers' methods to understand their students' funds of knowledge can be a valuable activity in a learning community. For example, the Michigan Assessment Consortium's<sup>3</sup> work on model assessment systems provides rich portraits of early literacy development and assessment that illustrates what it means to understand students' funds of knowledge (Michigan Department of Education, 2020). Such resources can act as a source of ideas for teachers. As they try out some of the suggested approaches, teachers can collectively share what they are learning about their students' funds of knowledge, develop local strategies to continue to deepen their knowledge of their students, and explore how they can make connections between students' home knowledge and experiences and the ideas they are learning about in school.

With information about students' funds of knowledge in hand, teachers can craft new learning goals or modify existing ones, as the teacher in Tables 5-1, 5-2, and 5-3 did by using Diego Rivera's work to reflect students' family backgrounds. Sharing learning goals with peers and discussing ways to incorporate students' funds of knowledge can strengthen teachers' skills in connecting learning goals to students' lived experiences, enhancing their motivation and their identity as capable learners. Reviewing students' responses after the lesson with colleagues can also provide insights into the motivational and identity aspects of the learning goals.

Leaders can also assist teachers in developing a deep understanding of the local community and students they are teaching. By spearheading discussions with their teacher colleagues, leaders can help them develop deeper community knowledge and think about how this knowledge can be used to positively impact learning goals, curriculum, teaching, and assessment.

## **Elicitation of Assessment Evidence**

### *Creating Formative Assessment Opportunities*

Ambitious teaching provides the means for designing formative assessment into teaching, particularly when teachers can draw on strong disciplinary knowledge. Teachers can review exemplar lesson plans and discuss with their learning community peers the ways in which multiple opportunities to elicit evidence are intentionally embedded throughout the lesson, how these elicitations align to the learning goals and success criteria, and what student responses are anticipated. Administrators or coaches could assist in locating these exemplar plans. Alternatively, a video recording of a lesson could be reviewed in a learning community to consider how evidence was elicited and acted on.

Teachers can collaboratively examine the ways that assessment evidence is generated and consider the extent to which important disciplinary ideas and practices are represented in assessment tasks or classroom discussions. For example, if a science assessment task for summative purposes provides lockstep directions for students to set up equipment and collect data, it will provide very few insights into students' understanding of science concepts or practices.

---

<sup>3</sup> See <https://www.michiganassessmentconsortium.org/elas>.

Teachers can also use and expand their disciplinary knowledge by collectively evaluating the degree to which the assessment task, whether for formative or summative purposes, enables students to show their thinking in multiple discipline-based modes. For instance, assessing a student's understanding of the main argument of a text might include matching different elements of the argument to specific paragraph numbers, a short answer response, and a graphic organizer to note the key points about two authors' views of the same topic.

Working on a peer's lesson plan can also be fruitful. This exercise may enhance formative assessment opportunities by considering how the lesson plan builds on students' prior knowledge and whether there are multiple entry points to allow all students to demonstrate what they know. A group of teachers could also review a lesson plan after the lesson has been taught in order to discuss if the evidence elicited was as they anticipated and provided sufficient and actionable insights into student learning. After this review, teachers might modify how they elicited evidence for this learning goal and related success criteria for future use.

### *Scaffolding Student Self-Assessment*

In a learning community or grade-level meeting, teachers can collaborate on strategies to support students' use of self-assessment. These strategies could include co-developing opportunities for self-assessment tied to particular lessons or generating questions for students to think about while they are involved in learning activities to support their metacognitive thinking. For example, in mathematics, students could be asked: "What is the problem about? What are the similarities and differences between the problem at hand and the problems you have solved in the past and why? What are the appropriate strategies, tactics, or principles for solving the problem and why? Does my solution make sense?" (Mevarech & Kramarski, 1997).

Teachers can also share protocols they use to support self-assessment or co-develop them for specific lessons. For instance, a middle school mathematics teacher uses the success criteria for the lesson with a Likert scale for them to rate their level of understanding and space for them to write something about what they learned or what they need to learn more about. A group of teachers developed a learning log for students to complete at the end of a class period that included questions such as: "What was successful about your learning today? What difficulties or problems did you encounter in your learning? How did you manage those difficulties?" In addition to supporting students' self-assessment and self-regulatory processes, student responses become important sources of evidence for teachers to use in planning next steps (Heritage, 2021).

The deliberate practice of trying out strategies, discussing how the strategies worked in a lesson with peers, making revisions, and trying them out again can help increase teacher expertise in scaffolding student self-assessment over time.

### *Reviewing Summative Assessments*

Teacher inquiry cycles, as a form of deliberate practice, provide a structure for repeated rounds of examination of teacher-created summative assessments. For example, a group, possibly in a learning community, could focus on a single assessment to follow the process from task design and selection to use, interpretation, and action. Collectively, the group could review the purpose of the assessment and how it will be communicated to students, whether students are adequately prepared for the assessment, and whether the assessment's design sufficiently matches the learning goals. After any needed revisions have been made, the assessment can be used. Afterwards, the group can reconvene to discuss how students responded to the assessment task, what insights into student learning were gained, and what future revisions might be needed. An inquiry cycle related to teacher-created summative assessment as well as required external summative assessments can be guided by questions such as those in Box 5-6.

The deliberate practice of repeating teacher inquiry cycles once a month or once per quarter will allow for the in-depth examination of an assessment task that will help teachers learn how to individually review other assessments they are using against the same criteria. For teachers who do not have colleagues using the same assessment tasks, peers can still serve as a sounding board, even if they do not have student data to compare across classes. Such a review can increase teachers' knowledge about the assessment and potentially lead to discarding or modifying the assessment to better serve summative needs.

### *Reviewing the Set of Assessments Within a Unit*

Examining the full set of formative and summative assessment tasks and prompts used within a unit is a worthwhile activity for individual teachers and teacher meetings to determine whether there is coherence between what is assessed for formative and summative purposes. Depending on the length of the unit and the number of tasks, this

#### **BOX 5-6** **Questions to Guide the Analysis of Summative Assessments**

- How well are students' funds of knowledge and interests represented across assessment items?
- Do the items have sufficient entry points to provide all students with opportunities to show where they are in their learning in ways that situate them as competent (e.g., open-ended problem solving tasks that are accessible to all students)?
- Are the items meaningful to students and will they perceive them as worthwhile?
- Do the items align with curricular and instructional goals (i.e., have the students had opportunities to learn what is represented on the items)?
- Do the items integrate cognitive, social and emotional, and cultural dimensions of learning?
- Will the assessment provide information that can be used to advance student learning?

process might need to be applied to a sample rather than all assessments. The opening review could analyze alignment with learning goals, whether there are opportunities for students to respond in multiple ways, if the assessments promote the major disciplinary ideas of the unit, and the extent to which the tasks draw on students' funds of knowledge and linguistic diversity. The review could also include student input, including their perceptions of the various assessment approaches and what they think their purposes are. Documenting what is being learned during the review can then inform future modifications to the tasks. After using the revised tasks, questions, and prompts, teachers can reflect on how the changes impacted student responses and the quality of the evidence generated.

### **Interpretation of Evidence and Action**

#### *Engaging in Evidentiary Reasoning and Interpretation*

As noted in this chapter's introduction, assessment is a process of reasoning from evidence. Building evidentiary reasoning skills is an essential component of becoming assessment literate and is dependent on disciplinary knowledge and interpretive skills. For instance, Jim Minstrell and colleagues' research shows how science teachers are able to make more nuanced interpretations of evidence when they possess strong disciplinary knowledge combined with high levels of interpretive skills (Minstrell et al., 2009). These teachers reasoned, for example, that their students expressed speed as proportional to the net force acting on the object, whereas less skilled teachers only noted that students were wrong about the net force needed (Minstrell et al., 2009). Learning progressions, whether research-based or locally developed, can help teachers learn what to notice in student responses, particularly when the progressions highlight common student misconceptions and less sophisticated ways of thinking.

Evidentiary reasoning and interpretation skills can also be developed by analyzing student work with associated rubrics, guided by questions like: "What and how are my students thinking in relation to the learning goal? What are the strengths of their thinking? What are the next steps for students to deepen their learning?" During a lesson, teachers who are skilled in formative assessment will have these questions in mind when they are observing students, asking questions, and listening to student talk so that they can interpret what the evidence they are obtaining reveals about learning. Discussions with colleagues after a lesson might consist of sharing the evidence observed in the student work and describing inferences the teacher made in real time. To evaluate student achievement beyond "got it" or "didn't get it," the same questions posed at the start of this paragraph can be answered when teachers review summative data, particularly if the data are accompanied by clear learning goals and success criteria.

Knowing a student well is also part of evidentiary reasoning and interpretation. For example, a teacher might infer that a student was drawing on their funds of knowledge as a basis for understanding a disciplinary concept, and subsequently leverage this knowledge in an asset-based way in determining next steps for the student. Similarly, when teachers have access to street data, their interpretations of a student's academic performance may be augmented by considering these data. For instance, teachers may be cognizant of economic challenges in the local community that can cause stress for

families or be knowledgeable about the traumas experienced by recent immigrants and how these traumas might impact learning and assessment performance. However, teachers will also need to be conscious of their own web of privileges and inequities, since they may shape their perceptions and influence their work with students. This consciousness is essential if teachers are to ensure fair and justice-oriented interpretations of street data. Accumulating this street data is not a one-time activity but is rather knowledge that teachers assemble over the course of a school year or longer (González, 2005), often with the assistance of school administrators.

### *Making Contingent Responses*

Often one of the most challenging skills for teachers to develop is taking contingent action based on interpreted evidence (Heritage et al., 2009). There are various ways, however, to support the development of this skill. For instance, in a learning community, one teacher might share lesson plans that integrate formative assessment, describe the evidence elicited to her peers, discuss the pedagogical action taken in response—for example, modeling, explaining, or prompting—and then evaluate how effective these responses were for advancing learning. Lesson revision suggestions to support improved formative assessment can also be discussed. Such exchanges can lead to shared lesson structures and routines, which can benefit students as they move between teachers within a grade, or from grade to grade.

Teachers could share lesson plans and explain how they leveraged student ideas or their funds of knowledge as bridges to disciplinary concepts. Teachers could also solicit ideas from colleagues about how to more effectively bridge to disciplinary concepts, with respect to specific interpretations of evidence. Discussions about students' funds of knowledge or ways in which students present ideas may conflict with some teachers' assumptions about students' families, backgrounds, or abilities. Again, developing one's sociocultural consciousness can help mitigate such assumptions. Mutual trust in professional learning situations, like teacher learning communities, will be paramount for surfacing and working through any conflicts in a supportive manner.

Asset-based interpretations of student learning and contingent responses can be augmented by teachers' knowledge and the application of an underlying learning progression (Wylie et al., 2018, p. 147). For example, recognizing that since a student cannot represent a proportional reasoning problem numerically but can describe the situation using "more than" and "less than" phrases, there is an opportunity to build on the student's initial understanding of the problem. Similarly, noticing that a student was able to laboriously solve a proportional reasoning problem using a build-up strategy but not able to use a more efficient scalar approach provides a starting point for a discussion of multiple solution strategies—rather than allowing a deficit mindset to simply see this student as having failed to use cross-multiplication (Wylie et al., 2018, p. 147).

Collaborative lesson or unit planning can help teachers build their repertoire of contingent actions by thinking together about possible student responses to specific activities and subsequent potential strategies to advance learning. In-the-moment formative assessment is especially supported through this planning process. A group video analysis of a lesson or a written vignette that integrates formative assessment



could be scaffolded by a peer or coach and focus on the contingent actions a teacher took, why they thought the teacher took the specific action, and how and why they thought it was or was not effective. If the video is of a teacher from the learning community, having a discussion with the teacher about the actions taken after the viewing and analysis presents an even further benefit.

### *Planning Feedback*

Skills in providing feedback to students as a type of contingent response, particularly in formative assessment, can be developed by sharing and critiquing examples of feedback from other teachers or from external sources like practitioner books. Deliberate practice in giving individual feedback on the same pieces of work and then sharing, discussing, and revising the feedback if needed can be an ongoing focus of a teacher learning community. Collaborative consideration of how students' funds of knowledge can be included in feedback can also be beneficial in strengthening the utility of that feedback, as can sharing examples of how teachers have provided feedback that draws from knowledge of their students. Examining how students have used feedback as a form of reverse engineering can also be a means of reviewing the quality and effectiveness of the provided feedback. Secondary school teachers who are teaching large numbers of students across several classes, in particular, can share strategies for how they manage to efficiently provide some form of feedback to all their students—for instance, comment markers linked to specific criteria on a specific piece of work.

Students also need to have feedback about their performance on summative assessments, which help them understand how well they met the goals of the unit or course and then assist them in setting goals for future learning. Teachers could discuss how they approach this within their learning community: Do they have one-on-one conferences with all students or just those they believe need extra support? Do they provide written comments to students about their performance, and then have students respond with their own perspective, or do they make a plan with students for how they will accomplish the goals they set during the next unit? If teachers are using data from summative assessments to assign grades, they might consider the learning benefits of providing feedback on summative assessments and then giving students the chance to revise their work or retake the assessment based on the feedback. When students have feedback about summative assessments and act on that feedback, the summative nature is temporary because teachers are using the data formatively and learning is still in progress (Brookhart, 2017).

### *Supporting Peer Feedback*

In addition to teacher feedback, peers can also assess each other's work and provide feedback to support revision, which is beneficial to both the giver and receiver (Rollinson, 2005; Spiller, 2012). Teachers can collaborate on strategies to teach students how to give feedback—for instance, sharing common ways of introducing students to peer assessment and feedback, discussing strong and weak examples of feedback with students, or conducting teacher think-alouds to demonstrate to students how they think

about feedback in relation to specific pieces of work. Teachers could share protocols they use for scaffolding peer feedback—for instance, an elementary teacher developed a feedback structure for her students to use when commenting on peers’ work: P (put up: what the student is doing well), Q (a clarifying question about the work), and S (a suggestion for improvement). When other teachers in her school heard her share how it was helping her students improve the quality of their peer feedback, they also started to use it (Heritage & Wylie, 2020). Teachers can also review ways that they invite students to respond to peer feedback and how students responded. Teachers can try out these strategies with their students and later debrief with their peers to learn from each other’s experiences, making revisions if needed. Beyond teaching strategies and support protocols, the success of peer feedback will very much depend on the classroom culture—another factor teachers need to keep in mind as they are collaborating on strengthening students’ peer feedback skills.

### *Using Classroom Assessment Data Evaluatively*

Teachers can collaboratively develop clear grading criteria for summative assessments that describe the quality of the desired student performance while avoiding compliance factors for work completion or following classroom procedures (Guskey & Brookhart, 2019). Examining student work with teacher colleagues against shared grading criteria can increase teachers’ interpretive skills and lead to more consistent grading. In the same vein, asking a colleague to review one’s grades can help inter-rater reliability and increase consistency among teacher grading practices.

Individually and collectively, teachers should critically examine their grading practices to identify bias, particularly related to students’ behavior and participation in the classroom (Taylor & Nolen, 2022). For instance, research on teacher judgments about student behavior suggests that teachers reprimand students of color more often than White students for subjective infractions in the classroom (Taylor & Nolen, 2022). Most instances of bias are unintentional, but taking a hard look at one’s own grading practices can help ensure more equitable grading.

Results from required external summative assessments can be used evaluatively to examine the effectiveness of curriculum materials and pedagogical approaches to inform future revision and use. Teachers can review the data using questions such as those suggested in Box 5-7. Districts often have specific protocols for examining summative data, including large-scale assessment results. These protocols can be useful resources for administrators or coaches to lead a review with teachers of summative data and collaboratively make improvement plans.

Conversations about summative data among teachers require making individual teaching approaches more transparent and a school- or department-wide culture of curiosity, grounded in the belief that any unit can be revised and taught better in the future. It is important for teachers to remember, and for administrators to reinforce, that they have control over what and how they teach, and that anything teachers can learn together about how to modify their practices will ultimately benefit their students.

### **BOX 5-7**

#### **Questions to Guide the Analysis of External Summative Assessment Data**

- Are there patterns in data that suggest one or more concepts were difficult for many students? If so, might adjustments to curriculum materials or pedagogical approaches for future use be warranted, in order to better support student understanding?
- What problems of practice do the data suggest? How can they be ameliorated?
- For students just assessed, are there opportunities in future units to revisit concepts that some were struggling with?
- Did students perform differently across classes on one or more concepts? If so, did teachers use different pedagogical strategies when teaching those concepts which might be useful for all students?
- Do subgroups of students perform differently on one or more concepts? If so, are there implications for how to engage all students in the learning?
- Are the same patterns of difference in performance visible in the unit level assessments? If so, what are areas to pinpoint for intervention?

## **SYSTEM SUPPORT FOR ASSESSMENT LITERACY**

In this section, we address how school and district leaders as well as state departments of education can provide systemwide support to promote assessment literacy.

### **School and District Leaders**

We have already noted specific ways that school and district leaders can establish a climate and community to strengthen assessment literacy among their teacher colleagues, as well as ways that they can contribute to specific assessment literacy learning opportunities. In this section of the chapter, we describe more general attitudes and skills that leaders should embody to effect change with respect to assessment literacy within their schools and districts. It bears emphasizing that school and district leaders should model and cultivate a culture of curiosity that supports productive failure and in which constructive assistance is offered to teachers whose first attempts to change practice may not be successful (Wylie et al., 2009; Youngs & King, 2002).

School and district leaders need to be familiar with the assessment literacy knowledge and skills described earlier in this chapter so that they can recognize the need for sustained professional learning for teachers and can participate in productive discussions about assessment literacy topics with their colleagues (Heritage et al., 2017). This knowledge base is essential for promoting high-quality assessment practices and bringing coherence among local priorities or mandates. For example, a leader needs to recognize and challenge when there are philosophical differences across policies—like a district-level mandate requiring a grade to be provided for every piece of student work—that are antithetical to school-based efforts to emphasize feedback rather than grading. Sometimes efforts to improve classroom assessment practices are directly undermined by other policies or implicit expectations regarding assessment. Pressures, real or perceived, to improve school or district performance on state assessments can

result in classroom practices that are more focused on test preparation than learning, exacerbating the dangers of narrowing curriculum to only focus on content that will be assessed on state-required assessments (Wylie & Gholson, 2023).

School and district leaders also need to ensure the coherence of, and then make evident relationships among, local initiatives like instructional reforms and standards implementation; formative assessment practices; and Diversity, Equity and Inclusion programs, and how assessment literacy pertains to each one. In doing so, leaders can promote more integrated approaches to teacher learning and help reduce teachers' feelings of being overwhelmed and frustrated by the perception of these initiatives being the "the flavor of the week" (cf. Fullan, 2010). In this regard, leaders could utilize a disciplined approach to inquiry, such as a Plan-Do-Study-Act cycle (Russell et al., 2020), which stimulates a common way of thinking about assessment that can align with other priorities, like those at the school, district, or state level. In the Plan-Do-Study-Act cycle, participants can collectively raise and explore questions about student learning with respect to local initiatives.

Leaders can establish clear policies for assessment use. Such policies may emphasize formative assessment practices and help teachers "right-size" the role of required external assessment data, including using them in ways that do not unduly and inappropriately dominate classroom pedagogical decisions. In instances where there are significantly different outcomes from classroom and external assessments, leaders should consider explanations for those differences with their teacher colleagues. For example, it may be that there is different content coverage or different parts of the standards emphasized on an external assessment than on classroom assessments, such as no or limited representation of mathematics practices or primarily selected-response items on external assessments. There may be different expectations of proficiency, different approaches for how students can display what they know and can do, or students may perceive the content as not worthwhile or regard the assessment as having no bearing on their experiences—particularly if there is an absence of cultural context in the assessment. Pursuing these explanations together can also serve a moderating function, supporting the development of consistent expectations across teachers within a school (Heritage & Wylie, 2020).

School and district leaders need to be creative problem solvers to identify time in already packed schedules for when teachers can collaborate. What works in one context may not be directly transferable to another, and leaders cannot assume that teachers will "make time" because this work is important.

Finally, leaders must be comfortable with ambiguity. There is no single correct place to start with assessment literacy. There is a logic and order to assessment literacy knowledge and skills, but local needs, teacher interests, and experience will suggest different starting points. Teacher agency and engagement in learning is often more important than having the "right" starting point.

## State Leaders

While state educational leaders have less direct contact with school leaders, teachers, and classrooms, they play an important role from a policy perspective and can set the context for how schools and districts perceive and use state accountability assessment and other data.

First, state leaders can provide explicit policy support for the reconceptualization of assessment to reflect cultural, social and emotional, and cognitive dimensions, and the value of classroom assessment in the service of ambitious teaching and learning. In this regard, it is essential that state policy use consistent language around components of an assessment system and language that aligns with research-based guidance for formative assessment (Wylie, 2022). While outside their jurisdiction, when state leaders treat formative assessment as more than just frequent summative assessment, which our own experience suggests they often do, they neglect its potential value to student learning and equity. State leaders need to be sensitive to the tension between using assessments for accountability and using assessments for teaching, learning, and development—and emphasize the value of latter purpose (Gordon et al., 2012).

Second, to lead the creation of a statewide culture for equitable and just assessment, state leaders will need to examine individual and systemic privileges to develop a socio-cultural consciousness that can both inform and permeate policy. State leaders must ensure that their policies are sensitive to the communities they serve and, in particular, those who have been historically marginalized or disenfranchised. For example, the Oregon Department of Education published a guidance document in 2021 that aligned six federal and state programs into a single planning document, and encouraged districts to apply an equity lens to their funding applications.<sup>4</sup> An equity lens is “an active tool that supports core values, commitments, orientations, and questions to become standard practice” (Oregon Department of Education, 2022, p. 37) and “applying an equity lens helps create a systematic structure and process to ensure that no focal group or community is ignored in the process of community engagement and plan development” (Oregon Department of Education, 2022, p. 97). Another example of modeling a focus on equity comes from the Wisconsin Department of Public Instruction, which published the Model to Inform Culturally Responsive Practices to support educators in developing the beliefs, knowledge, and practices needed to meet the needs of all Wisconsin students (Wisconsin Department of Public Instruction, 2017).

Third, state adoptions of curriculum materials need to be equity-focused, and, in the case of local-control states, clear guidance about relative strengths and weaknesses of curricula should be made available to support district or school decisions (Polikoff, 2021). An equity-focused curriculum that integrates cognitive, social and emotional, and cultural dimensions of learning and that promotes a culturally sustaining pedagogy will provide the bedrock for assessment use and for the development of assessment literacy among educators at all levels.

Fourth, state leaders can provide high-quality professional learning materials and supports for assessment literacy—for example, learning progressions, exemplars of learning goals, videos of practice, coaches, or other kinds of expert support that are sustained over time. The Michigan Department of Education is a case in point: in

---

<sup>4</sup> See [https://www.oregon.gov/ode/StudentSuccess/Documents/ODE\\_Integrated%20Guidance](https://www.oregon.gov/ode/StudentSuccess/Documents/ODE_Integrated%20Guidance).



collaboration with the Michigan Assessment Consortium, the state board of education endorsed Assessment Literacy Standards (Michigan Assessment Consortium, 2015, 2017, 2020), which are supported by a broad array of programs and services, including the Assessment Learning Network.<sup>5</sup> The Formative Assessment for Michigan Educators, now in its 15th year, is a statewide program offering sustained professional learning in formative assessment, which has been implemented widely across the state.<sup>6</sup>

Finally, state leaders can also serve as conveners to support work being done at the district level (P. Leonard, personal communication, February 2023). State leaders serving as conveners can take the form of bringing in external national or local experts, facilitating district-to-district sharing on an issue that is relevant to all, or creating university partnerships. States and school districts often have access to data that needs analyzing, and universities have graduate students looking for opportunities to apply what they are learning in measurement or evaluation programs. One example of this work has been led by the Connecticut State Department of Education, resulting in the creation of the Centre for Connecticut Education Research Collaboration.<sup>7</sup> The Center for Connecticut Education Research Collaboration currently has relationships with 11 public and private universities across the state, and are engaged in a wide variety of studies aimed at supporting Connecticut educators and students (A. Gopalakrishnan, personal communication, May 2023).

### KEY TAKEAWAYS

Assessment literacy is critical so that teachers can equitably use assessment in the service of student learning and development. The body of knowledge and skills required to be assessment literate is extensive but should not be thought of as work for individual teachers to tackle in isolation. Rather, with collaboration among teachers and the appropriate local- and state-level support for teacher learning, it is eminently achievable (see Box 5-8 for key ideas for assessment literacy and professional learning).

---

<sup>5</sup> See <https://www.michiganassessmentconsortium.org/aln>.

<sup>6</sup> See <https://famemichigan.org>.

<sup>7</sup> See <https://portal.ct.gov/ccerc>.

**BOX 5-8**  
**Key Ideas for Assessment Literacy and Professional Learning to Support Just and Equitable Teaching**

- Classroom assessment should draw on a combination of formative and summative practices, and ambitious teaching provides a rich and reciprocal context for formative assessment in particular.
- Whether the assessment purpose is formative or summative, integration of learning goals that address cognitive, social and emotional, and cultural components of learning; assessment evidence; and asset-based and future-oriented interpretations and actions are necessary, and all aspects should take place within a safe and trusting classroom climate.
- Teachers need access to a strong, coherent, and standards-aligned core curriculum that provides the necessary backbone for both teaching and assessment.
- Each aspect of assessment—learning goals, assessment evidence, and asset-based and future-oriented interpretations and actions—must be considered through a lens of equity and justice.
- There is a professional body of knowledge and skills that is required for assessment literacy, and it should include the examination of individual and systemic privileges to develop schoolwide sociocultural consciousness.
- School and district leaders play a critical role in establishing a climate that supports teacher learning, inquiry, and practice, while state leaders can create a policy environment that supports classroom assessment.

**REFERENCES**

- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*(3), 261–271. <https://doi.org/10.1037/0022-0663.84.3.261>.
- Andrade, H. L., & Heritage, M. (2017). *Using formative assessment to enhance learning, achievement, and academic self-regulation*. Routledge. <https://doi.org/10.4324/9781315623856>.
- Au, K. H., & Kawakami, A. J. (1994). Cultural congruence in instruction. In E. R. Hollins, J. E. King, & W. C. Hayman (Eds.), *Teaching diverse populations: Formulating a knowledge base* (pp. 5–23). State University of New York Press.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117–148. [http://dx.doi.org/10.1207/s15326985ep2802\\_3](http://dx.doi.org/10.1207/s15326985ep2802_3).
- Bang, M., & Medin, D. (2010). Cultural processes in science education: Supporting the navigation of multiple epistemologies. *Science Education, 94*(6), 1008–1026. <https://doi.org/10.1002/sce.20392>.
- Bennett, R. E. (2019). Integrating measurement principles into formative assessment. In H. L. Andrade, R. E. Bennett, & G. J. Cizek (Eds.), *Handbook of formative assessment in the disciplines* (pp. 20–33). Routledge.
- Bergan, J. R., Sladeczek, I. E., Schwartz, R. D., & Smith, A. N. (1991). Effects of a measurement and planning system on kindergartners' cognitive development and educational programming. *American Educational Research Journal, 28*(3), 683–714. <https://doi.org/10.3102/00028312028003683>.
- Black, P. J. (1993). Formative and summative assessment by teachers. *Studies in Science Education, 21*(1), 49–97. <https://doi.org/10.1080/03057269308560014>.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. <http://dx.doi.org/10.1080/0969595980050102>.
- Brookhart, S. M. (2017). *How to give effective feedback to your students* (2nd ed.). ASCD.
- Brown, G. T. L., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research, 3*(1), 22–30. <https://doi.org/10.14786/flr.v2i1.24>.

- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698–712. <https://doi.org/10.1080/02602938.2012.691462>.
- Calvert, L. (2016). The power of teacher agency: Why we must transform professional learning so that it really supports educator learning. *The Learning Professional*, 37(2), 51–56. <https://learningforward.org/wp-content/uploads/2016/04/the-power-of-teacher-agency-april16.pdf>.
- Cowie, B., Harrison, C., & Willis, J. (2018). Supporting teacher responsiveness in assessment for learning through disciplined noticing. *The Curriculum Journal*, 29(4), 464–478. <https://doi.org/10.1080/09585176.2018.1481442>.
- Darling-Hammond, L., Hyster, M. E., Gardner, M., & Espinoza, D. (2017). *Effective teacher professional development*. Learning Policy Institute. [https://learningpolicyinstitute.org/sites/default/files/product-files/Effective\\_Teacher\\_Professional\\_Development\\_REPORT.pdf](https://learningpolicyinstitute.org/sites/default/files/product-files/Effective_Teacher_Professional_Development_REPORT.pdf).
- Duijnhouwer, H., Prins, F. J., & Stokking, K. M. (2010). Progress feedback effects on students' writing mastery goal, self-efficacy beliefs, and performance. *Educational Research and Evaluation*, 16(1), 53–74. <https://doi.org/10.1080/13803611003711393>.
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363–406. <https://doi.org/10.1037/0033-295X.100.3.363>.
- Esteban-Guitart, M., & Moll, L. C. (2014). Lived experience, funds of identity and education. *Culture & Psychology*, 20(1), 70–81. <https://doi.org/10.1177/1354067X13515940>.
- Feldman, J. (2019). *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms*. Corwin.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal*, 28(3), 617–641. <https://doi.org/10.2307/1163151>.
- Fullan, M. (2010). The big ideas behind whole system reform. *Education Canada*, 50(3). Retrieved on 11.1.2023 from <http://michaelfullan.ca/wp-content/uploads/2016/06/13396082070.pdf>
- Goe, L., Wylie, E. C., Bosso, D., & Olson, D. (2017). State of the states' teacher evaluation and support systems: A perspective from exemplary teachers. *ETS Research Report Series 2017(1)*, 1–27. <https://doi.org/10.1002/ets2.12156>.
- González, N. (2005). Beyond culture: the hybridity of funds of knowledge. In N. González, L. C. Moll, & C. Amanti (Eds.), *Funds of knowledge: Theorizing practices in households, communities, and classrooms*. Lawrence Erlbaum Associates.
- Gordon, E. W., Gordon, E. W., Aber, L., & Berliner, D. (2012). Changing paradigms for education: From filling buckets to lighting fires to cultivation of intellectual competence. *The Gordon Commission on the Future of Assessment in Education*. [https://www.ets.org/Media/Research/pdf/gordon\\_gordon\\_berliner\\_aber\\_changing\\_paradigms\\_education.pdf](https://www.ets.org/Media/Research/pdf/gordon_gordon_berliner_aber_changing_paradigms_education.pdf).
- Guskey, T. R. & Brookhart, S. M. (2019). *What we know about grading: What works, what doesn't, and what's next*. ASCD.
- Gutiérrez, K. D., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, 32(5), 19–25. <https://doi.org/10.3102/0013189X03200501>.
- Heritage, M. (2021). *Formative assessment: Making it happen in the classroom* (2nd ed.). Corwin.
- Heritage, M., Gerzon, N., & Mancuso, M. (2017). *Necessary conditions for successful implementation of formative assessment*. WestEd. <https://www.gettingsmart.com/wp-content/uploads/2017/09/Necessary-Conditions-for-successful-implementation-of-Formative-Assessment-June-18-DRAFT.pdf>.
- Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment. *Educational Measurement: Issues and Practice*, 28(3), 24–31. <https://doi.org/10.1111/j.1745-3992.2009.00151.x>.
- Heritage, M., and Wylie, E. C. (2020). *Formative assessment in the disciplines: Framing a continuum of professional learning*. Harvard Education Press.
- Levinson, C. & Brantley-Netwon, V. (2017). *The youngest marcher: The story of Audrey Faye Hendricks, a young civil rights activist*. Atheneum Books for Young Readers.
- Mevarech, Z. R., & Kramarski, B. (1997). Improve: A multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal*, 34(2), 365–394. <https://doi.org/10.3102/00028312034002365>.

- Michigan Assessment Consortium. (2020). *Assessment literacy standards*. [https://www.michiganassessmentconsortium.org/wp-content/uploads/MAC\\_AssessLitStds\\_2017\\_9.19.17.pdf](https://www.michiganassessmentconsortium.org/wp-content/uploads/MAC_AssessLitStds_2017_9.19.17.pdf).
- Michigan Department of Education. (2020). *Early literacy assessment systems that support learning*. [https://www.michiganassessmentconsortium.org/wp-content/uploads/ELAS\\_Guide\\_2020\\_ToPrint.pdf](https://www.michiganassessmentconsortium.org/wp-content/uploads/ELAS_Guide_2020_ToPrint.pdf).
- Minstrell, J., Anderson, R., & Li, M. (2009). *Assessing teacher competency in formative assessment: Annual report*. University of Washington: FACET Innovations.
- Mislevy, R. J. (1994). Evidence and inference in educational measurement. *Psychometrika*, 59(4), 439–483. <https://doi.org/10.1007/BF02294388>.
- Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement*, 33(4), 379–416. <https://doi.org/10.1111/j.1745-3984.1996.tb00498.x>.
- Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory Into Practice*, 31(2), 132–141. <https://doi.org/10.1080/00405849209543534>.
- Nasir, N. S., Lee, C. D., Pea, R., & McKinney de Royston, M. (2021). Rethinking learning: What the interdisciplinary science tells us. *Educational Researcher*, 50(8), 557–565. <https://doi.org/10.3102/0013189X21104725>.
- National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. The National Academies Press. <https://doi.org/10.17226/24783>.
- National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. National Academy Press. <https://doi.org/10.17226/10019>.
- Oregon Department of Education. (2022). Aligning for student success: Integrated guidance for six ODE initiatives. [https://www.oregon.gov/ode/StudentSuccess/Documents/ODE\\_Integrated%20Guidance.pdf](https://www.oregon.gov/ode/StudentSuccess/Documents/ODE_Integrated%20Guidance.pdf).
- Panadero, E., Jonsson, A., & Strijbos, J.-W. (2016). Scaffolding self-regulated learning through self-assessment and peer assessment: Guidelines for classroom implementation. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 311–326). Springer International Publishing.
- Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher*, 41(3), 93–97. <https://doi.org/10.3102/0013189X12441244>.
- Paris, D., & Alim, H. S. (Eds.). (2017). *Culturally sustaining pedagogies: Teaching and learning for justice in a changing world*. Teachers College Press.
- Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist*, 36(2), 89–101. [https://doi.org/10.1207/S15326985EP3602\\_4](https://doi.org/10.1207/S15326985EP3602_4).
- Pellegrino, J. W. (2014). Assessment as a positive influence on 21st century teaching and learning: A systems approach to progress. *Psicología Educativa*, 20(2), 65–77. <https://doi.org/10.1016/j.pse.2014.11.002>.
- Pellegrino, J. W., Grover, S., Harris, C. J., & Wiebe, E. (2023). Classroom assessment in STEM education: An introduction to the report. In C. J. Harris, E. Wiebe, S. Grover, and J. W. Pellegrino (Eds.), *Classroom-based STEM assessment: Contemporary issues and perspectives* (pp. 4–20). Community for Advancing Discovery Research in Education. <https://cadrek12.org/sites/default/files/2023-08/Classroom-Based%20STEM%20Assessment%20FULL%20REPORT.pdf>.
- Polikoff, M. (2021). *Beyond standards: The fragmentation of education governance and the promise of curriculum reform*. Harvard Education Press.
- Pryor, J. (2010). Pedagogies of in/equity: Formative assessment/assessment for learning. In H. H. Krüger, U. Rabe-Kleberg, R. T. Kramer, & J. Budde (Eds.), *Bildungsungleichheit revisited*. VS Verlag für Sozialwissenschaften. [https://doi.org/10.1007/978-3-531-92201-0\\_8](https://doi.org/10.1007/978-3-531-92201-0_8).
- Randall, J. (2021). “Color-neutral” is not a thing: Redefining construct definition and representation through a justice-oriented critical antiracist lens. *Educational Measurement: Issues and Practice*, 40(4), 82–90. <https://doi.org/10.1111/emip.12429>.
- Randall, J., Slomp, D., Poe, M., & Oliveri, M. E. (2022). Disrupting white supremacy in assessment: Toward a justice-oriented, antiracist validity framework. *Educational Assessment*, 27(2), 170–178. <https://doi.org/10.1080/10627197.2022.2042682>.
- Raza, A., Penuel, W. R., Li, J., & Frank, K. A. (under review). Enhancing the relevance of science learning for students owed an education debt: Exploring the potential of the Student Experience Inquiry Cycle. *Journal of Research in Science Teaching*.

- Rollinson, P. (2005). Using peer feedback in the ESL writing class. *ELT Journal*, 59(1), 23–30. <https://doi.org/10.1093/elt/cci003>.
- Russell, J. L., Correnti, R., Stein, M. K., Bill, V., Hannan, M., Schwartz, N., Booker, L. N., Pratt, N. R., & Matthis, C. (2020). Learning from adaptation to support instructional improvement at scale: Understanding coach adaptation in the TN mathematics coaching project. *American Educational Research Journal*, 57(1), 148–187. <https://doi.org/10.3102/0002831219854050>.
- Safir, S., & Dugan, J. (2021). *Street data: A next-generation model for equity, pedagogy, and school transformation*. Corwin.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. <https://doi.org/10.3102/0013189X02900700>.
- Shepard, L. A. (2019). Classroom assessment to support teaching and learning. *The ANNALS of the American Academy of Political and Social Science*, 683(1), 183–200.
- Shepard, L. A. (2021). Ambitious teaching and equitable assessment: A vision for prioritizing learning, not testing. *American Educator*. <https://files.eric.ed.gov/fulltext/EJ1321974.pdf>.
- Sherin, M., Jacobs, V., & Philipp, R. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. Routledge.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23.
- Sireci, S. G. (2020). Standardization and UNDERSTANDARDIZATION in educational assessment. *Educational Measurement: Issues and Practice*, 39(3), 100–105. <https://doi.org/10.1111/emip.12377>.
- Spiller, D. (2012). *Assessment matters: Self-assessment and peer assessment*. The University of Waikato.
- Stembridge, A. (2020). *Culturally responsive education in the classroom: An equity framework for pedagogy*. Routledge.
- Stiggins, R., & Duke, D. (2008). Effective instructional leadership requires assessment leadership. *Phi Delta Kappan*, 90(4), 285–291. [http://www.pdkmembers.org/members\\_online/publications/archive/pdf/k0812sti.pdf](http://www.pdkmembers.org/members_online/publications/archive/pdf/k0812sti.pdf).
- Stronge, J. H., & Xu, X. (2021). *Qualities of effective principals* (2nd ed.). ASCD.
- Subero, D., Vila, I. & Esteban-Guitart, M. (2015). Some contemporary forms of the funds of knowledge approach. Developing culturally responsive pedagogy for social justice. *International Journal of Educational Psychology*, 4(1), 33–53. <http://dx.doi.org/10.4471/ijep.2015.02>.
- Taylor, C. S., & Nolen, S. B. (2022). *Culturally and socially responsible assessment: Theory, research, and practice*. Teachers College Press.
- Toom, A., Pyhältö, K., & O'Connell Rust, F. (2015). Teachers' professional agency in contradictory times. *Teachers and Teaching: Theory and Practice*, 21(6), 615–623. <https://doi.org/10.1080/13540602.2015.1044334>.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. McGraw-Hill Education.
- Villegas, A. M., & Lucas, T. (2001). *Educating culturally responsive teachers: A coherent approach*. State University of New York Press, Albany.
- Walqui, A., Hartman, L., Schmida, M., & Feldman, S. (2023). *English language arts units for English learners*. National Research and Development Center to Improve Education for Secondary English Learners.
- Wenger, E., McDermott, R., & Snyder, W. M. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Harvard Business School Press.
- Wisconsin Department of Public Instruction. (2017). *Equity: Wisconsin's model to inform culturally responsive practices*. <https://dpi.wi.gov/sites/default/files/imce/rti/pdf/mlss-wi-model-inform-crp.pdf>.
- Wylie, C. (2022). If I ruled the world ... Three suggestions for state department leaders to increase formative assessment practices in classrooms. *Center for Assessment*. <https://www.nciea.org/blog/if-i-ruled-the-world>.
- Wylie, C., Bauer, M., Bailey, A. L., & Heritage, M. (2018). Formative assessment of mathematics and language: Applying companion learning progressions to reveal greater insights to teachers. In A. L. Bailey, C. A. Maher, & L. C. Wilkinson (Eds.), *Language, literacy, and learning in the STEM disciplines: How language counts for English learners*. Routledge Taylor & Francis Group.



- Wylie, E. C. & Gholson, M. (2023). U.S. federal assessment policy reforms in the twenty first century and their impacts. *International Encyclopedia of Education* (4th ed.) (pp. 108–118). <https://doi.org/10.1016/B978-0-12-818630-5.09027-8>.
- Wylie, E. C., & Lyon, C. J. (2016). *Using the formative assessment rubrics, reflection and observation tools to support professional reflection on practice*. Council of Chief State School Officers. [https://center.ncsu.edu/ncfalcon/pluginfile.php/2/course/section/57/FAROP\\_Revised\\_2016.pdf](https://center.ncsu.edu/ncfalcon/pluginfile.php/2/course/section/57/FAROP_Revised_2016.pdf).
- Wylie, E. C., & Lyon, C. J. (2020). Developing a formative assessment protocol to support professional growth. *Educational Assessment*, 25(4), 314–330. <https://doi.org/10.1080/10627197.2020.1766956>.
- Wylie, E. C., Lyon, C. J., & Goe, L. (2009). Teacher professional development focused on formative assessment: Changing teachers, changing schools. *ETS Research Report Series*, 2009(1), i–32. <https://doi.org/10.1002/j.2333-8504.2009.tb02167.x>.
- Youngs, P., & King, M. B. (2002). Principal leadership for professional development to build school capacity. *Educational Administration Quarterly*, 38(5), 643–670. <https://doi.org/10.1177/0013161X02239642>.

# District and School Practices and Assessments to Support a Learning-Centered Vision

*Elena Diaz-Bilello, University of Colorado Boulder*

*Jared Anthony, Carnegie Foundation for the Advancement of Teaching*

## CONTENTS

INTRODUCTION .....	168
THE ROLE OF THE DISTRICT.....	170
A Culture of Testing .....	171
ASSESSMENTS TO SUPPORT AMBITIOUS TEACHING AND LEARNING .....	172
An Example of Classroom Assessment Supporting Ambitious Teaching .....	174
Distinguishing the Role of Distal Assessments.....	176
Establishing School Partnerships .....	177
ESTABLISHING AN INSTRUCTIONAL INFRASTRUCTURE FOR AMBITIOUS TEACHING .....	178
Access to High-Quality Curriculum .....	179
Professional Learning.....	184
Grading .....	188
TAKING A <i>LEARNING SYSTEMS</i> APPROACH TO IMPLEMENTING A LEARNING-CENTERED VISION .....	192
CONCLUSION .....	193
REFERENCES.....	194

## INTRODUCTION<sup>1</sup>

In 2017, *Education Week* reported that in a nationally representative sample of more than 500 U.S. K–12 teachers, approximately 85 percent indicated that they had experienced new changes or reforms in the past two years, and more than 58 percent indicated that they were experiencing “reform fatigue” (Loewus, 2017). Most teacher respondents (85 percent) further shared that “as soon as they get a handle on a new reform, it changes” (Loewus, 2017). In Chapter 8 of this volume, “Developing, Implementing, and Institutionalizing Complex Educational Innovations: Considerations for Balanced Assessment Systems,” Peurach and Linn frame this persistent state of churning reforms as part of the conventional narrative that “policy-level fragmentation, incoherence, and turbulence” foster the same within local districts, as their central offices and schools attended to changing policy ambitions and priorities.

Based on the perspectives from educators and educational researchers outlined in *Education Week*, it should come as little surprise that incoherence also characterizes the design of assessment systems in many districts and schools. For example, the website of one large urban school district in a Western state shows the district’s “balanced assessment system framework” as a large menu of assessments that fall under either Assessments for Learning (formative) or Assessments of Learning (summative). Under the formative category, there is a list of more than 12 assessments, including universal screeners and district-mandated interim assessments. Under the summative category, there is a large list of district-required and state-required year-end tests. The website notes that multiple types of assessments and data from multiple occasions are needed to guide instruction and improve student performance. In other words: it seems that the simple act of selecting and administering assessments under both categories defines balance in this assessment system without consideration for whether data from all of these assessments are communicating a coherent picture of student performance to effectively inform instructional steps.

Unfortunately, the misconception that a balanced assessment system means using several different types of assessments is also perpetuated by some vendors that sell interim assessments. As Marion (2021) notes, several test vendors claim that a balanced assessment system should consist of a selection of assessments—formative, interim, and summative—that teachers and administrators can combine to form a comprehensive picture of student learning. This conception of a balanced assessment system will likely lead to a patchwork of assessments that do not advance a particular vision or model of learning and often inspires a refrain heard in the education field about districts and schools being “data rich but information poor.”

In this chapter, we discuss the practices and structures employed by districts and schools functioning as *learning systems* (see Chapter 8 of this volume, “Developing, Implementing, and Institutionalizing Complex Educational Innovations: Consider-

---

<sup>1</sup> We are so grateful to our reviewers (Amy Berman, Debbie Durrence, Peter Leonard, Jonathan Supovitz, and Scott Marion) for providing us with thoughtful feedback that challenged and extended our thinking. We also want to thank the school districts that we work with for the incredible partnership opportunities. We could not have written this chapter without learning from our work with you. And finally, we cannot thank Lorrie Shepard enough, as she always and generously made the time to share her insights and provide us with feedback throughout this project.

ations for Balanced Assessment Systems,” for a fuller discussion). We argue that these practices and strategies can be used to support and sustain assessments that focus on ambitious teaching and learning. Districts and schools characterized as being in the *learning systems* stage of the systems continuum that includes *school systems*, *education systems*, and *learning systems* are “distinguished by capabilities to engage diverse stakeholders ... in collaborating to develop the shared understandings, knowledge, and values needed to identify and address local educational ambitions, needs, and problems” (Chapter 8 in this volume, p. 259). Thus, districts and schools in the *learning systems* stage are best positioned to implement a learning-centered vision as addressed throughout this volume. *Learning systems* stage institutions demonstrate the importance of building collaborative networks between districts, schools, and other key partners to tackle important problems and issues, such as addressing inequities in the education system.

This chapter opens with a brief account of how school districts have recently influenced the instructional work taking place in U.S. schools. This accounting shows that districts do not have a long legacy of engaging with schools as *learning systems* and demonstrates that several districts remain in what Peurach and Russell (in Chapter 8 of this volume, “Developing, Implementing, and Institutionalizing Complex Educational Innovations: Considerations for Balanced Assessment Systems”) refer to as the *education systems* and *school systems* stages. Districts with *school systems* characteristics focus on supporting the business and administrative functions of operating schools that serve their communities and tend to more weakly support improving schools’ educational work. Districts with *education systems* characteristics, on the other hand, focus on improving teaching and learning, but their goals often involve achieving technical effectiveness and efficiency in response to federal and state policy goals and interventions. Balanced assessment systems operating in districts located in either of these two stages would not resemble the type of instructional and assessment work that supports ambitious teaching and learning practices, described in great detail by Ruiz-Primo and Furtak in Chapter 4 of this volume, “Classroom Activity Systems to Support Ambitious Teaching and Assessment.”

We then consider what it would look like for a *learning system* district to use assessments supporting ambitious teaching and learning. Under this scenario, districts would prioritize the use of classroom assessments and use federal- and state-required test results in ways that do not detract and instead support effective classroom assessment practices. We walk through an example of classroom assessments that support this learning-centered vision, and in so doing clarify ideal features and qualities of these assessments.

We note, as Kang et al. (2016) documented in their study of teachers who use rich instructional tasks, that having access to rich tasks and assessments does not necessarily mean that they will be implemented effectively—the successful enactment of rich classroom tasks and assessments requires a strong supportive infrastructure. Thus, we discuss *instructional infrastructure* and provide examples of ambitious teaching and learning practices and structures that district and school personnel located in a *learn-*

*ing system* stage should support and invest in (Hopkins & Spillane, 2015; Hopkins & Woulfin, 2015). *Instructional infrastructure* components discussed in order of priority are:

- high-quality curricula,
- professional learning, and
- grading.

Although we discuss each component separately, the three should work together to establish an assessment system integrated with instruction. The *instructional infrastructure* literature commonly distinguishes assessment as a separate component of teaching that supports instruction, but we do not do so, because, as discussed by Ruiz-Primo and Furtak in Chapter 4 of this volume, “Classroom Activity Systems to Support Ambitious Teaching and Assessment,” the types of assessment that support ambitious teaching and balanced assessment are inseparable from instruction.

Lastly, we address how districts might begin working with schools to move toward this bold vision for teaching and learning while simultaneously engaging in necessary evaluations to monitor implementation. Because districts and schools can be located anywhere along the *school systems, education systems, or learning systems* continuum, we provide only general ideas for how districts and schools can begin this highly complex work aimed at changing organizational behaviors, cultures, policies, and structures.

## THE ROLE OF THE DISTRICT

In most school districts in the United States today, central office leadership sets the vision and policies for improving teaching and learning (Honig & Coburn, 2008). Districts communicate budget priorities, provide instructional frameworks and curricular materials, set expectations for assessment strategies, and provide professional development opportunities for school-based personnel (Coburn et al., 2009; Honig & Venkateswaran, 2012; Penuel et al., 2017).

This has not always been the case. Historically, district personnel tended to focus on fulfilling business and compliance functions rather than implementing a vision of teaching and learning for their schools (Honig, 2013). Peurach and Russell in Chapter 8 of this volume, “Developing, Implementing, and Institutionalizing Complex Educational Innovations: Considerations for Balanced Assessment Systems,” classify districts that focus on business and compliance as being in a *school systems* stage, in which district management concern themselves with the structural and procedural activities needed to deliver educational services.

Districts first began to play a more active role in teaching and learning in the 1980s, when research highlighted the important role that they could and should play in fostering effective schools (Mac Iver & Farley, 2003). Over the subsequent two decades, such research generated momentum for researchers and policymakers to consider how districts could prominently steer instructional reforms (Honig & Coburn, 2008). The advent of the federal No Child Left Behind Act (NCLB) of 2001 further cemented districts’ direct involvement in steering the teaching and learning visions for their schools, particularly because NCLB provided monetary incentives to encourage districts to take a leading role in evidence-based school improvement work (Anderson & Young, 2018; Leithwood et al., 2019; No Child Left Behind Act, 2001).



The passage of NCLB not only encouraged many districts to take a prominent role in improving teaching and learning in their schools but was also the starting point for these districts to focus substantial resources and energy on raising test scores to improve school accountability ratings (Au, 2007; Blazar & Pollard, 2017). In shifting their resources and attention to these types of educational work, many districts entered what Peurach and Russell in Chapter 8 of this volume, “Developing, Implementing, and Institutionalizing Complex Educational Innovations: Considerations for Balanced Assessment Systems,” classify as the *education system* stage. By establishing new structures and organizational practices to directly influence education work, district personnel became directly involved with instruction to meet or respond to federal and state policy goals, including implementing academic content standards and accountability policies for student outcomes in mathematics and English language arts (e.g., Marsh, 2002; Massell & Goertz, 2002; Snipes et al., 2002). The implications of this type of organizational structure, with its intention to improve educational practices, opened the door to the testing-focused quandary discussed next.

### A Culture of Testing

Despite NCLB’s well-intentioned efforts to focus on the performance of minoritized student groups, many authors point to how district and school practices and policies shaped by state testing and accountability largely exacerbated rather than mitigated inequalities (e.g., Au, 2007; Blazar & Pollard, 2017; Hamilton et al., 2008). These practices included restructuring or narrowing curricula to focus instructional activities on content for state tests, adjustments to programming and scheduling (e.g., removing art classes from the school’s program or reducing or removing recess time) to better prepare students for the tests, and using highly scripted curricula and strict pacing to help improve test scores (Au, 2007; Blazar & Pollard, 2017; Crocco & Costigan, 2007; Dresser, 2012; Duncan-Owens, 2009; Hamilton et al., 2008; Heiser et al., 2015).

NCLB also motivated districts to make large investments in commercially developed interim assessments designed to efficiently monitor student learning and collect predictive information about student performance on high-stakes summative state tests (Shepard, 2017). Despite persistent calls from multiple stakeholders for reduced state testing (Olson & Jerald, 2020), at present, many school districts continue to administer a large array of tests to students, as highlighted in Figure 6-1.

The areas in Figure 6-1—showing a Grade 11 calendar during the 2020–2021 school year—highlighted in yellow mark time in which different groups of students are scheduled to be taken out of their classrooms to participate in formal district interim testing or summative state testing. Most days of most months are earmarked for some sort of testing, underscoring how testing continues to shape scheduling and behaviors at many U.S. districts and schools.

During the NCLB period (2001–2015), many educational researchers documented how learning experiences for students of color and other minoritized groups were negatively impacted by testing and accountability (Blazar & Pollard, 2017; Heiser et al., 2015; Teoh et al., 2014). Today, researchers still note the persistence of such practices within a testing culture and have documented how they continue to harm minoritized groups (Gitomer & Iwatani, 2022; Randall et al., 2022). The call Marion and colleagues



**FIGURE 6-1** Testing calendar from an urban high school.  
 NOTE: This photo was shared with the authors of this chapter on the condition of not revealing the school or district location.

make in Chapter 1 of this volume, “Reimagining Balanced Assessment Systems: An Introduction,” to re-center the focus of assessments on classroom assessment, not only represents a demand to reprioritize the types of assessments used by districts and schools, but is a call to action for states, districts, and schools to ensure that this shift affords greater equity and fairness in the educational opportunities offered to students.

### ASSESSMENTS TO SUPPORT AMBITIOUS TEACHING AND LEARNING

Because the learning-centered vision articulated in this volume is informed by sociocultural learning theory (see Chapter 3 of this volume, “Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems”) and plays an integral part in ongoing teaching and learning activities (see Chapter 1 of this volume, “Reimagining Balanced Assessment Systems: An Introduction”), district and school leaders will want to prioritize classroom assessments that are situated in *classroom activity systems*. Such systems are “largely determined by a teacher’s pedagogical actions [and] provide affordances for participation in a community of practice” (Kang & Furtak, 2021, p. 75). According to Kang and Furtak (2021), a *classroom activity system* can be identified by the set of interactions—or relational work—inspired by activities involving the materials and structures deployed in classrooms (e.g., lesson plans, assessments, instructional tasks, instructional routines, etc.). Kang and Furtak (2021) further note that the degree to which learners participate in these activities depends on “who they are and their historical relationship with the discipline, and actors [i.e., other students, teachers] in classrooms” (p. 75). This implies that when the materials and structures used in a *classroom activity system* are intentionally designed to improve the quality

of relational interactions between participants, the participation of all students in the classroom can be optimized. Since proximal and near classroom assessments represent the major focus of learning-centered assessment systems, forms of assessment that are distal from a *classroom activity system*, such as state tests and other locally required assessments, have more utility for district and school administrators than for educators and students (Ruiz-Primo et al., 2012).

Forms of classroom assessment that support this dynamic and relational *classroom activity system* enable districts to play a highly influential role in supporting ambitious teaching and learning practices. That is, a district can serve as a learning hub for schools by providing school leaders and educators with professional learning and resources to support the selection, development, and use of classroom assessment as an integral part of high-quality curriculum and instruction. Taken together, the formative and summative classroom assessments discussed in Chapter 4 of this volume, “Classroom Activity Systems to Support Ambitious Teaching and Assessment,” form a coherent and relatively comprehensive picture of student learning. Importantly, the specifications for the design of such assessments lean heavily on processes that not only reveal student thinking and reasoning but also help foster a trusting and inclusive learning environment for all.

Such classroom assessments described in Chapter 4 of this volume, “Classroom Activity Systems to Support Ambitious Teaching and Assessment,” are connected to the learning environment because assessments that support rich and culturally responsive instructional and learning opportunities can generate observable classroom practices such as teachers and students co-constructing knowledge as they explore ideas in depth; respectful dialogue that values the ideas of every learner; and teachers avoiding “front-loading” vocabulary in classroom interactions to signal that there is not one “correct” way to use language in a given discipline (Darling-Hammond & Cook-Harvey, 2018; Suárez et al., 2020; Thompson et al., 2021; Windschitl et al., 2012, 2018). These practices indicate that an important feature of a classroom with a culture of ambitious teaching is that it positions students to collaborate with teachers in the learning and assessment experiences enacted in the classroom.

Performance assessments that can be used in multiple ways (e.g., end-of-unit assessments, capstone performance demonstrations, or common district-wide assessments) are a natural fit for this ambitious teaching vision. Such assessments engage students in complex tasks and activities—synthesizing information, evaluating evidence, problem-solving—and also provide relatively accurate markers of the higher-order skills and knowledge students have acquired (Conley, 2015; Darling-Hammond & Adamson, 2014; Faxon-Mills et al., 2013; Hofman et al. 2015; Linn & Burton, 1994). In fact, a wide range of informal and formal formative assessment strategies and processes can be used to elicit student thinking in line with this vision. These include discourse-based strategies employing extended discussions to discover how students are thinking about the topic at hand and then adjusting daily instruction accordingly. Such assignments are instrumental in enacting ambitious teaching. We next walk through an example of the types of formative and summative assessments that can be used in a *classroom activity system* and how the assessments work together to support ambitious teaching and learning.

## An Example of Classroom Assessment Supporting Ambitious Teaching

To illustrate how formative and summative classroom assessments can cohere and produce a rich body of information to help inform teaching and learning, we point, as an example, to the curriculum development and assessment work of the Storylines Project, based at Northwestern University. This project seeks to advance the implementation of Next Generation Science Standards (NGSS) in districts and schools by providing high-quality, open-resource materials. The Storylines Project exemplifies a coherent approach for using science assessments in service of curricular and instructional goals since the curriculum materials or units of study developed for different grades has embedded informal and formal assessments that support the learning targets connected to big disciplinary ideas in science, and ultimately support the enactment of NGSS standards.

According to the Storylines project team, each open-source unit with lessons and embedded assessments presents “a coherent sequence of lessons in which each step is driven by students’ questions that arise from their interactions with phenomena” (Next Generation Science Storylines, n.d.). In other words, students serve as key collaborators in the learning process, helping to move the classroom forward by explaining scientific phenomena or solving problems. By positioning students as active participants in their learning, Storylines units attend to key features of sociocultural learning theory and can help educators implement ambitious teaching in their classrooms.

Each Storylines unit developed is intended to elicit the intentional enactment of formative instructional strategies and tasks from teachers and help surface student reasoning and inquiry. Teachers can then use rich summative performance assessments to evaluate student learning at the end of the instructional period or unit of study. Importantly, teachers can use these assessment experiences as part of their lessons. For example, one portion of the set of curricular resources provided for a Grade 4 unit was titled, “*Why do some things wash up on the beach and others don’t?*” Informal checks of collaborative group work produced for one of the seven lessons include having teachers:

Look at student responses in the “Make Predictions” and “Make Plans” sections on page 1 of Student Handout 4.1 to see students’ ideas about how to create the type of waves they need, and how to consistently carry out their plans. Also check page 2 of the handout to see if students are able to accurately record their data from multiple trials. (Aycock et al., 2019)

If group work products indicated that students were struggling with the tasks assigned for this lesson, the unit encourages teachers to draw on formative strategies as follows:

If students are not able to summarize this thinking about how waves move floating objects, have them look back at the “Finding Patterns” section on page 3 of Student Handout 4.1. Ask them what was similar among most or all of the group’s data. If students are struggling to comprehend that the results were not what they predicted, remind them that science is often surprising, and unexpected results help us ask more questions and design better investigations next time. It may also be helpful to remind students that the “How we represent our thinking” section of their Progress Tracker

can be done with drawings in addition to words - some students may be more able to explain their thinking with a labeled model than complete sentences, and that's okay! (Aycock et al., 2019)

The intention of these informal instructional and assessment strategies is to encourage students to build on prior knowledge to advance their thinking. The strategies also encourage students to consider different modalities for demonstrating knowledge and skills and to take risks with their investigations.

Once teachers reach the end of a Storylines unit, they may opt to administer a summative unit test that allows students to formally demonstrate their learning. Figure 6-2 presents a student's response to a single item on the summative test for the Grade 4 unit described above. This rich item attends to the performance expectations evaluated throughout the unit but also asks students to demonstrate the depth of their learning by transferring the acquired knowledge and skills to an entirely new set of tasks.

**1A. What will you do to make the most, tallest waves you can? Use the words disturbance, amplitude, and wavelength to describe how you will push the water to make the most, tallest waves.**

To make taller waves, you need to make a ~~big~~ disturbance. Making a faster, bigger disturbance will make high ~~amplitude~~ amplitude, and short wave length, that will make a lot of high waves to hit your friend.

**1B. Draw a model of what your waves would look like in the water. Be sure to label your model including the words amplitude, and wavelength.**

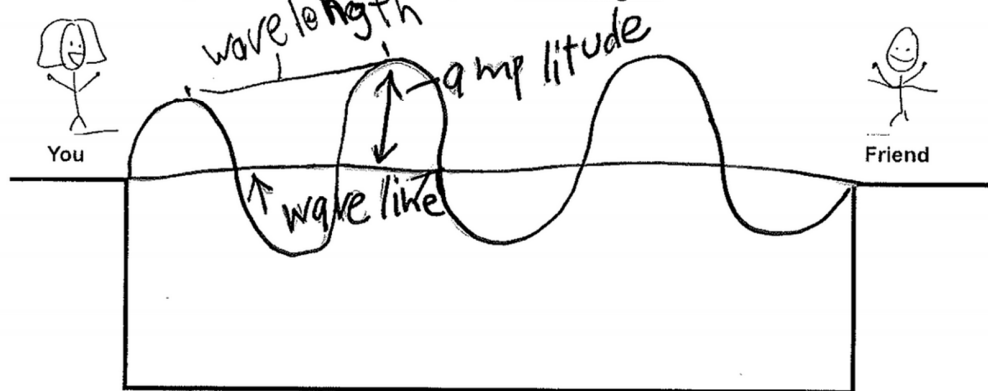


FIGURE 6-2 Example of a rich curriculum-embedded item on a Grade 4 end-of-unit summative test. SOURCE: Aycock et al. (2019).



Storylines units developed for science educators provide an example of how a district and school can achieve *horizontal coherence*<sup>2</sup> in assessments supporting one disciplinary area (science) using a variety of informal and formal curriculum-embedded assessments that accompany each lesson. These lessons achieve *horizontal coherence* because the set of assessment experiences and opportunities embedded in each unit address important learning targets linked to the disciplinary “big ideas” established in the curriculum (Shepard et al., 2018).

Taken together, the set of rich formative and summative assessment experiences provided in the curriculum units developed by the Storylines Project give teachers a clear picture of what students know and can do relative to key learning goals deliberately aligned with the performance expectations set by academic standards. The Storylines Project also works with a conception of coherence that requires teachers to factor in student perspectives and agency as key design principles for developing instructional and assessment experiences. The project’s developers believe that learning and assessment can only gain “coherence” when students participate in both activities as co-constructors of knowledge (Reiser et al., 2021). Thus, in the Storylines Project, student perspectives and agency operate as important defining features for establishing coherence in the set of learning and assessment experiences offered under the banner of ambitious teaching and learning.

### **Distinguishing the Role of Distal Assessments**

Distal assessments—assessments that are external to classroom learning activities—can play an important role in helping school and district leaders evaluate broader school-level performance. However, these assessments serve a distinct purpose from classroom assessments. Because federal and state education agencies recognize that districts exercise direct oversight over schools, districts are charged with ensuring that required state and federal assessments are administered to students. This oversight function requires districts to review student results from standardized assessments because such assessments generate comparable information about schools, allowing district and school leaders to identify important student performance trends relative to key school and student initiatives, reforms, and other interventions.

Other locally required assessments, such as interim tests classified as formative classroom assessments by many district and school leaders and educators, would actually be considered distal assessments. Although some disagree with classifying these district-developed or commercially developed interim tests as distal assessments (e.g., Dyer, 2017), these assessments are not designed to support the culturally responsive practices and relevant learning experiences embedded in a *classroom activity system* as described in Chapter 4 of this volume, “Classroom Activity Systems to Support Ambi-

---

<sup>2</sup> To clarify this term, we refer back to Chapter 1 of this volume, “Reimagining Balanced Assessment Systems: An Introduction”: “At the classroom level, coherence generally means ensuring that assessments are consistent with high-quality curricula and instructional materials that reflect contemporary understandings of disciplinary learning and knowledge development. *Horizontal coherence* is alignment among curriculum, instruction, and assessment to help students develop proficiency in a content domain” (National Research Council, 2006). . . . *Horizontal coherence* is most critical at the classroom level, especially because formative and other classroom assessments must cohere with ambitious instruction and an equity-centered curriculum. School districts generally have the authority to support *horizontally coherent* systems of assessment since curriculum and other related decisions are generally made at the district level” (p. 5, italic in the original).

tious Instruction and Assessment.” Given that interim tests tend to rely on a selected response format, they provide limited opportunities for making student thinking and reasoning visible. Thus, these tests are not ideal for a classroom assessment system that seeks to deepen teaching and learning (Perie et al., 2009; Shepard, 2019).

Districts will want to clearly communicate to school leaders that distal assessments are not part of a thriving *classroom activity system*. While the results of distal assessments are essential for broader program evaluation and district-wide monitoring of schools, they should not drive classroom instruction. Ensuring that districts message these priorities to schools is consistent with state practices that take up two of the high-leverage state actions highlighted in Chapter 7 of this volume, “State Practices and Balanced Assessment Systems,” clearly communicating the role of state summative assessments and mitigating their misuse and the misuse of other locally required tests. In the next section, we address the infrastructure practices districts and schools should adapt to support and sustain balanced assessment systems.

### Establishing School Partnerships

An important first step for districts seeking to encourage ambitious teaching is partnering with schools to strengthen investment in this vision of teaching and learning. As mentioned in the previous section, this type of collaboration includes clearly communicating which assessments should be prioritized and why. Recognizing that some schools (e.g., charter and innovation-zone schools) are autonomous from their districts and can shape their own educational goals and vision and that some districts have adopted decentralized structures (i.e., more authority and resources shift to schools), this chapter addresses a common scenario found in many states where districts guide and motivate the operations and performance of schools—district leadership setting the vision for teaching and learning and central offices expected to coordinate with schools to implement this vision.

Even under this scenario, schools do not necessarily follow the district’s strategic direction and vision, and some schools face difficulties enacting desired reforms. Factors that can engender difficulties include the size of the school district; school-based leaders and personnel who misinterpret the vision and accompanying policies; internal conflicts within the district; and the extent to which the district’s existing organizational structures, policies, and norms obstruct school-based initiatives (Massell & Goertz, 2002; Snipes et al., 2002; Togneri & Anderson, 2003). Literature on school turnaround is rife with case studies that document the failure of reforms to take root in schools, particularly when districts enforce top-down implementation (Meyers, 2020).

The idea that schools can effectively improve teaching and learning on their own, without resources and guidance from their districts, their state, or both, is not supported by the available evidence (Honig & Rainey, 2015; McLaughlin & Talbert, 2002; Polikoff, 2021). Even so, most researchers recognize that districts need to provide schools with the flexibility to take up proposed reforms in ways that best fit their needs and contexts (Elmore & Burney, 1997; Marsh, 2002; Massell & Goertz, 2002; McLaughlin & Talbert, 2002; Meyers, 2020; Togneri & Anderson, 2003). That is, to have school personnel buy into a district’s vision, school leaders must have the flexibility to make decisions about how to engage in proposed changes since schools have different levels of readiness to

take up complex reforms. At the district level, flexibility entails paying attention to the localized context of each school and the communities it serves to determine what level of support and resources are appropriate on a site-by-site basis (Massell & Goertz, 2002; Meyers, 2020).

In addition to exercising flexibility, districts can establish reciprocal relationships that benefit both districts and schools. To do this, they can look to strong partnership models such as the research–practitioner partnership model (Coburn et al., 2013b). Coburn et al. (2013b) define the research–practitioner partnership model as a long-term collaborative relationship established between researchers and practitioners to attend to persistent issues or problems. By encouraging schools to take up productive adaptations of the support and resources provided by central offices, districts can enable schools to identify and implement actions or steps that would best facilitate their adoption of the district’s teaching and learning vision. This type of partnership work will likely promote schools’ sustained support of the district’s vision. We now turn to a discussion of *instructional infrastructure* in considering how ambitious teaching can be supported through partnerships between districts and schools.

### ESTABLISHING AN INSTRUCTIONAL INFRASTRUCTURE FOR AMBITIOUS TEACHING

In this section, we outline examples of district and school practices and policies to support three critical components of *instructional infrastructure*—high-quality curricula, professional development, and grading—that can catalyze both district and school efforts to recenter their focus on classroom assessments. Ideally, the state would partner with school districts to support classroom assessment, in a similar way that we would envision districts to partner with their schools (see Chapter 7 of this volume, “State Practices and Balanced Assessment Systems”). That is, districts can sustain their work to provide high-quality curricula and support high-quality professional learning opportunities at schools through the funding, resources, and other support they receive from the state. This partnership model would also be taken up by schools, so that school personnel could partner with parents, students, and other community members to improve the school’s *instructional infrastructure*.

Districts and schools seeking to implement this learning-centered vision are likely functioning as *learning systems* (see Chapter 8 of this volume, “Developing, Implementing, and Institutionalizing Complex Educational Innovations: Considerations for Balanced Assessment Systems”). As indicated in the introduction to this chapter, districts located in the *learning system* phase actively seek broad stakeholder input when embarking on and learning from this transformational work. Leaders at the district and school level whose institutions are at the *learning system* stage typically engage in interactive and relational practices with stakeholders to facilitate trust and buy-in from their stakeholders to deepen reforms by learning from stakeholder experiences and feedback. By learning from and through leaders’ multiple layered interactions with their broader network of stakeholders, these leaders can target their reform efforts. A key part of this broader relational work is to codesign the infrastructure—policies, structures, and practices—with the stakeholders who will support this complex instructional and assessment work in schools.

*Instructional infrastructure*, or what Penuel (2019) refers to as “infrastructuring,” includes components that contribute to the successful adoption of educational reforms, including how assessments are used. Specifically, *instructional infrastructure* includes the components aimed at shifting instructional practices and the set of interactions that occur within this infrastructure (Cohen et al., 2013; Mehta & Fine, 2015; Spillane et al., 2011). According to Hopkins and Spillane (2015), *instructional infrastructure* “forms a system intended, by design or default, to guide and monitor instruction and its improvement” (p. 422). Components within this system include but are not limited to, professional learning, assessments, instructional materials, instructional frameworks, school and district-level policies, roles, and positions focused on instructional support, programming, and oversight (Cohen et al., 2013; Hopkins & Spillane, 2015; Spillane et al., 2011). In theory, these components should work together to foster rich interactions in the *classroom activity system*.

If the components of an *instructional infrastructure* cohere to support a *classroom activity system* with ambitious teaching and learning as the organizing design, there is promise that the infrastructure will lead to desired outcomes like establishing an equity- and learning-centered environment for all students (Bryk et al., 2009). However, building such infrastructure will only gain traction in schools if district personnel engage with school leaders, educators, and other important stakeholders (e.g., community members) in a meaningful partnership.

Our discussion of *instructional infrastructure* begins with practices and policies that support the development of a high-quality curriculum since the presence of such a curriculum is a critical lever for “establishing coherent, consistent high-quality instruction in ... schools” (Polikoff, 2021, p. 103). In other words, ambitious teaching and learning—including classroom assessment practices—cannot happen unless schools have access to a high-quality curriculum.

Next, we shift the discussion to professional learning, as practices included in this component of *instructional infrastructure* enable schools and teachers to enact the curriculum and the assessments that support this infrastructure. We then address grading, as this component provides teachers and schools with an additional avenue for providing feedback that has the potential to improve learning. We do not define instruction as a separate component of *instructional infrastructure* because the above components combine to directly support instructional routines, including assessment. We acknowledge that additional infrastructural components could be examined—such as talent, career development, or teacher evaluations—but have limited the discussion to these three areas because they are critical levers for creating assessments that support a *classroom activity system* focused on ambitious teaching and learning.

### **Access to High-Quality Curriculum**

We open this section with an illustrative vignette sourced from personal conversations with Peter Leonard about the current work underway at Chicago Public Schools (CPS) to advance high-quality curriculum. This vignette (see Box 6-1) is intended to illustrate how a large school district has made significant investments in an *instructional infrastructure* to help spread this work in schools.

**BOX 6-1**  
**Chicago Public Schools: Advancing High-Quality Curriculum**

In 2020, the Chicago Public Schools (CPS) embarked on a reform to provide a high-quality curriculum (Skyline curricula) to all schools. This curriculum was a key component of their vision to advance student equity. The district made the decision to prioritize the provision of this curriculum and accompanying high-quality instructional materials and resources to all schools in order to ensure that all students, regardless of the school they attended, would have access to engaging and rigorous lessons and instructional materials. CPS also moved in this direction to signal a sweeping change to their vast network of schools: the district was turning away from using tests as the primary means for organizing instructional priorities and evaluating students, and would instead focus on improving the quality of teaching and learning provided to them. This meant that the district's assessment focus would prioritize curriculum-embedded classroom assessment that supported teachers in enacting the district's high-quality curriculum.

To engage in this work, the district established a broad group of curriculum experts—both within and external to CPS. Their task was to clearly define curricula in all content areas and grades, to select high-quality materials, and to work with curriculum partners to ensure that the selected materials were culturally responsive and relevant to Chicago students. In addition to defining the curriculum and materials, the district hired curriculum specialists in every disciplinary area and at every grade level to serve as professional learning partners for all schools throughout the district. Knowing that this initiative would not gain traction if schools were mandated to adopt the curriculum and materials for each disciplinary area, CPS showcased this work as a model that schools could choose to follow or to directly adopt (or not). The district's hope was that schools would see the quality of the investments it had made in this instructional infrastructure, and would be motivated to shift toward adopting the curriculum and the extensive resources the district had provided. Although the district is still in the early stages of elevating this instructional infrastructure, it has expended vast resources to implement it. Presently, 470 schools in the district use Skyline curricula in at least one grade band and content area. The support that central administrators and staff at CPS have given to this large endeavor communicates a unified front: that every stakeholder in the system prioritizes giving students opportunities to learn and benefit from a high-quality curriculum.

A high-quality curriculum—one that includes instructional frameworks, curriculum maps, instructional materials (inclusive of classroom assessment), and programming decisions—is the most critical component in a district or school's *instructional infrastructure*. This is because curriculum determines both the materials and resources that will directly support instructional activities, as well as the progression or sequencing of disciplinary content and skills teachers will use. The curriculum also determines when and which assessments should be used and how to organize the school schedule to maximize instructional time. In the following subsections, we focus on two specific curricular components needed to support an ambitious teaching vision: curriculum coherence and curriculum materials.

*Curriculum Coherence*

“In the absence of a learning plan with clear goals, how likely is it that students will develop shared understandings on which future lessons might build?” (Wiggins



& McTighe, 2005, p. 21). This question gets to the heart of why instructional frameworks and curricular maps are critical tools for communicating instructional guidance and identifying important areas of assessments. Because these frameworks and maps provide teachers with guidance and content for instruction, they can also facilitate coherence in the curriculum across and within grades by articulating expectations for each disciplinary area of the *instructional infrastructure* (Cohen et al., 2013; Hopkins & Spillane, 2015). As the authors of Chapter 4 of this volume, “Classroom Activity Systems to Support Ambitious Teaching and Assessment,” discuss, districts and schools can gain a clear understanding of the evidence needed to evaluate learning relative to established learning goals if they have a clear roadmap that outlines expectations for what should be taught in each disciplinary area. This understanding, in turn, should lead districts and schools to intentionally design or select assessments that match the learning goals, such as those used in the Storylines Project units.

Instructional frameworks and curriculum maps can help teachers specify assessment tasks and clarify which instructional moves should follow those tasks. When anchored to ambitious teaching as an organizational design, these frameworks and maps should also encourage teachers to attend to the developmental needs of students while simultaneously providing lessons that feature “well-scaffolded instruction and ongoing formative assessment(s) that support conceptual understanding, take students’ prior knowledge and experiences into account, and provide the right amount of challenge and support on relevant and engaging learning tasks” (Darling-Hammond et al., 2020, p. 98). As is noted in Chapter 4 of this volume, “Classroom Activity Systems to Support Ambitious Teaching and Assessment,” curricula designed for ambitious teaching not only respond to but also sustain the knowledge, practices, cultures, and languages of learners.

Teachers can also reference instructional frameworks and curriculum maps as guides in the development of instructional materials—such as units of study—and the accompanying set of curriculum-embedded assessments used to evaluate student performance. When teachers closely align their instructional and assessment practices to an instructional framework or curriculum map, this can be promising to establish more equitable access to high-quality learning opportunities for all students, regardless of the school they attend.

An example of an instructional framework designed to connect with ambitious teaching and learning is Schoenfeld’s Teaching for Robust Understanding (TRU) framework for math (Schoenfeld, 2013, 2017). The TRU framework consists of five dimensions, each of which focuses on what students are expected to do in math within the context of the learning activities enacted by teachers. These five dimensions are described in Figure 6-3.

Burkhardt and Shoenfeld (2019) explain how the five dimensions of the TRU framework can work together as principles for designing instruction connected to a well-specified sequence of learning activities and tasks that utilize formative assessment strategies. For example, Burkhardt and Shoenfeld (2019) note that, by using the TRU framework, teachers can design lessons that “uncover students’ existing ways of thinking, then create cognitive conflicts or disturbances that lead students to realize and confront inconsistencies ... through student-student and student-teacher discussion, in pairs or small groups, and then across the class as a whole” (p. 51). Teachers

The Five Dimensions of Powerful Classrooms				
The Content	Cognitive Demand	Equitable Access to Content	Agency, Ownership, and Identity	Formative Assessment
<i>The extent to which classroom activity structures provide opportunities for students to become knowledgeable, flexible, and resourceful disciplinary thinkers. Discussions are focused and coherent, providing opportunities to learn disciplinary ideas, techniques, and perspectives, make connections, and develop productive disciplinary habits of mind.</i>	<i>The extent to which students have opportunities to grapple with and make sense of important disciplinary ideas and their use. Students learn best when they are challenged in ways that provide room and support for growth, with task difficulty ranging from moderate to demanding. The level of challenge should be conducive to what has been called “productive struggle.”</i>	<i>The extent to which classroom activity structures invite and support the active engagement of all of the students in the classroom with the core disciplinary content being addressed by the class. Classrooms in which a small number of students get most of the “air time” are not equitable, no matter how rich the content: all students need to be involved in meaningful ways.</i>	<i>The extent to which students are provided opportunities to “walk the walk and talk the talk” – to contribute to conversations about disciplinary ideas, to build on others’ ideas and have others build on theirs – in ways that contribute to their development of agency (the willingness to engage), their ownership over the content, and the development of positive identities as thinkers and learners.</i>	<i>The extent to which classroom activities elicit student thinking and subsequent interactions respond to those ideas, building on productive beginnings and addressing emerging misunderstandings. Powerful instruction “meets students where they are” and gives them opportunities to deepen their understandings.</i>

FIGURE 6-3 The Teaching for Robust Understanding framework dimensions.  
SOURCE: Teaching for Robust Understanding Framework (n.d.).

using the TRU framework would then be asked to reflect on their formative practices and check whether they support rich mathematical content, achieve high levels of cognitive demand by maintaining productive struggles with content, ensure meaningful engagement for all students, and strengthen opportunities for student sense-making that fosters agency and identity (Burkhardt & Schoenfeld, 2019). Thus, implementing an instructional framework that supports ambitious teaching practices can also directly impact student academic performance. According to the authors, “Classrooms that did well on the rubric [connected to the TRU framework] did well on mathematics [classroom-based] measures ... [whereas] classrooms that scored poorly did not” (Burkhardt & Schoenfeld, 2019, p. 41).

Building on this example, we argue that if districts provided schools with resources to adopt this type of instructional framework for mathematics, it would set the groundwork for curriculum specialists to design and develop instructional and assessment strategies and routines. At the district level, district-based curriculum specialists—who are typically separated from district assessment staff—would be encouraged by leadership to work together to create educative resources and models that schools could directly adopt or reference in their selection of frameworks and materials. These curriculum specialists would be tasked with building capacity to support professional development work provided by the district and backed by school leadership.

District leaders who support this vision should recognize that offices supporting different functions will have to collaborate closely to establish coherence across the work of personnel who are charged with influencing teaching and learning. For example, in CPS, the curriculum, instruction, and assessment functions all reside within the Office of Teaching and Learning, which creates the structural conditions for closer collaboration between offices at the district level. This type of collaboration is rarely encountered in school districts (Latham, 2018). Nevertheless, the example is an important proof that departmental shifts can happen if they are prioritized by district leadership, and staff are provided the authority and resources to implement organizational change.

### *Curriculum Materials*

High-quality curriculum materials, inclusive of classroom assessments, should ideally reflect the content and activities specified in related instructional frameworks and curricular maps. We define “high-quality materials” as those that embody the learning-centered features described at length in previous chapters of this volume and the broader literature (e.g., Armstrong, 2021; Ladson-Billings, 2014; Reiser et al., 2021; Wang et al., 2021). Primary among these features is that the materials connect to the diverse experiences and interests of students. This feature is important because it has direct implications for designing learning and assessment experiences. Kaufman et al. (2020) note that, despite the importance of high-quality curriculum materials and resources, they are often lacking even though a few states have tried to help districts identify and invest in such materials. The disconnect between the high-quality curriculum materials promoted by these states and lower-quality materials can be attributed to the common policy of “local control,” under which curricular decisions are made by the district rather than by the state.

As addressed earlier in this chapter, each district should have a vision for its curriculum, but this vision is not always shared by schools. This is particularly the case when the district is large and oversees several networks of schools, as seen in the CPS vignette. Districts operating at the same scale as CPS understand that their vision for a high-quality curriculum is more likely to gain traction if it is promoted using strategies that build buy-in with schools. One aspect of the CPS strategy was to ensure that school leaders and educators were involved in the process of piloting and refining these high-quality curriculum models and resources so that they would be motivated to adopt these materials and resources. According to a CPS leader, this included working with school leaders and educators to formalize a clear definition of high-quality curriculum, evaluating current curricula along that quality definition, performing a non-evaluative curriculum audit in collaboration with schools and networks, and setting a multi-year goal toward high-quality curricula that empowers school leaders to lead that process in their buildings (P. Leonard, personal communication, 2023).

Another important strategy used by CPS to facilitate the adoption of high-quality materials in schools was to ensure that all educators had access to curriculum-specific professional learning from district-based content experts in order to effectively use these resources. These content experts, located in the district’s curriculum office, provided school-based staff with resources, learning opportunities, and additional support for implementing new curricular resources. Deploying strategies like those used by

CPS can lower some of the anticipated barriers and anxieties that can surface in schools when reforms, including new curricular and assessment directions, are implemented.

Polikoff (2021) recommends that states recruit teachers to participate in curriculum reviews to help and endorse high quality curriculum materials, as this is likely to improve buy-in for teachers to use those materials in the classroom. This strategy can be used by districts and schools that are in the process of adopting new curriculum materials and can also include the broader community, which would help ensure that the materials are responsive to the diverse backgrounds and interests of the local community. Another strategy is to have teachers participate in a curriculum adaptation of materials to better align existing materials with ambitious teaching and learning goals (Cook-Endres et al., 2014). Allowing teachers to engage in this type of adaptation work by reworking existing materials may be a more viable strategy in some schools and districts, particularly if a site does not have sufficient resources to supplant existing materials. This strategy would enable districts and schools to work with educators and modify selected units, tasks, and assessments to better align with learning-centered approaches and instructional models.

As discussed earlier in this chapter, districts can offer flexibility to schools in selecting instructional materials, especially when schools favor distinct programmatic and focal areas like expeditionary learning or a STEAM model (in which science, technology, engineering, and math are the focus, with the arts infused). However, flexibility will need to be balanced with the assurance that selected materials align with the high-quality criteria endorsed by the district in consultation with schools and other stakeholders (e.g., community-based groups). These criteria are critical for setting clear expectations about the features the instructional materials should embody—including embedded assessments—when individual schools are allowed to consider the materials they wish to select or adopt.

### **Professional Learning**

We begin this section with a vignette as an illustrative example of one school and district collaboration to support professional learning intended to elevate ambitious teaching practices. This vignette (see Box 6-2) summarizes key highlights from an existing report (Diaz-Bilello et al., 2022).

In the prior section, we touched on aspects of curriculum—coherence and materials—that provide the necessary support for important classroom-based instructional and assessment activities. Here, we note that these resources require instructional coaches, leaders, and teachers who have the knowledge and capacity to enact and use them in skillful ways. In other words, ensuring that all schools can access a high-quality curriculum is not enough—this work must be accompanied by personnel capacity building. To advance equity and create learner-centered instruction and assessment routines, both district and school leaders must provide teachers with professional development that enables them to build their knowledge and skills in the cognitive, social-emotional, and cultural dimensions of learning (see Chapter 5 of this volume, “Assessment Literacy and Professional Learning”).

Unless teachers build a repertoire of pedagogical content, knowledge, and skills to successfully enact a high-quality curriculum—which must include instructional and



**BOX 6-2**  
**Prairie Heights Middle School: Collaborative Opportunities in Professional Learning**

The Prairie Heights Middle School in Greeley, Colorado, participated in a network of turnaround schools established by the state and supported by their school district. Because the school had already invested in high-quality instructional materials approved by district and state partners, the school's leadership team turned its attention to revamping its professional learning structures and practices to improve their instruction and assessment practices. An important aspect of this restructuring work entailed ensuring that professional learning offered collaborative opportunities for teachers to try out and learn from the important formative instructional strategies used to engage students in their learning.

At Prairie Heights, all teachers review student work together. Every week in both grade-level and disciplinary-specific teams, they discuss instructional strategies and consider the home-life and personal circumstances of students. Coaching cycles were established so that school leaders and teacher leaders could mentor and provide feedback to novice teachers or teachers new to the school. The cycles were an opportunity for less experienced teachers to learn how to provide meaningful and actionable feedback to students. When teachers within these professional learning communities (PLCs) discuss the personal experiences of students, they use the understandings they have gained from having engaged with students and their families. This knowledge enables them to consider the issues and contexts that influence the learning experiences of students. School administrators serve as facilitators of these PLCs, along with other mentor teacher leaders. Together, the PLC and coaching cycles enable teachers to develop and test strategies—including building classroom structures intended to foster student engagement and collaboration—and receive feedback from mentors on how to refine the enacted routines. In the words of one teacher who had joined the school during the difficult turnaround period, “If it wasn't for the administrators and mentor teachers doing feedback loops, sticking with me, showing me what I'm missing, [and] pairing me with other teachers who could model for me, I probably would not have stayed.... I did not leave the school because of all of these supports to make me significantly better.”

assessment materials and resources—take-up of ambitious teaching vision will likely be less effective (Darling-Hammond, 2004; Darling-Hammond et al., 2017; Kang et al., 2016). For this reason, districts should work with school leaders to find creative ways of establishing and planning professional learning opportunities as a regular, ongoing, and frequent part of the school week (Darling-Hammond et al., 2017; Penuel et al., 2017, 2020a). For example, the Boulder Valley School District (BVSD) in Colorado recently enacted a policy of allowing their schools to set a late-start day once per week to allow the morning to be used for professional development. In an interview, a BVSD principal indicated that school leaders were “so grateful for this policy because [prior to this policy], we had so little time during the school year to spare on professional development” (S. Minnich, personal communication, April 25, 2023).

Districts that support this important infrastructure work can be highly influential in providing a policy framework and establishing clear expectations for school leaders in cultivating strong professional learning communities. In the opening vignette for this section, the Greeley district did just this in supporting Prairie Heights Middle School's decision to implement weekly horizontal and vertical team meetings for all teachers.



Even though each school ultimately defines the structure and culture of the professional learning it enacts for and with teachers, Kraft et al. (2021) note that the sense of success teachers have as professionals can be bolstered when they feel that they can depend on their district and school leaders to clearly communicate instructional priorities and provide them with targeted and relevant training.

### *Learning Communities and Distributed Teams*

Effective professional learning occurs in the context of a learning community—or culture—in which teachers engage in collective sense-making and knowledge construction (Coburn et al., 2013a; Fullan & Quinn, 2015; Hargreaves, 2000; Watson, 2014). Under the vision of ambitious teaching and learning outlined throughout this volume, such a learning community would ideally regularly convene to examine and discuss student work produced through rich classroom assessments and tasks and identify the instructional and assessment work needed to challenge, scaffold, and improve student learning. Districts can play an influential role in creating such learning communities, primarily by encouraging schools to build distributed leadership teams that empower teacher leaders to take on important mentorship roles. Central office staff can also provide schools with guidance, tools, and resources to support these professional learning communities. For example, furnish school leaders with protocols and tools to guide the analysis of student work produced in response to instructional tasks and assessments, and capture outcomes from these discussions. The district can also help model assessment focused work in professional learning communities by organizing the analysis of student work across educators located in different schools. One model to consider for facilitating the assessment focused work of professional learning communities is the twice-per-year student work analysis event that takes place across participating schools in the New York Performance Standards Consortium (Willis et al., 2022). The Consortium, which focuses on supporting curriculum-embedded, performance-based assessments, uses student work analysis to calibrate expectations and refine both the tasks and scoring rubrics used across participating schools.

However, it is important to note that the reach of the district is limited within schools and the efficacy of these structures depends on how school leaders define and organize their distributed leadership teams. When teacher leaders are provided with “resources, support, structure, and the [formal] authority” by school leaders, they can help drive school improvement efforts (Supovitz, 2018). Distributed teams can inform the selections of school curricula and/or assessments in their districts and provide effective coaching and expert support to challenge and transform common classroom practices that run counter to the vision of teaching, learning, and assessment their districts propose (Darling-Hammond et al., 2017; Wang et al., 2021). Even so, if the distributed leadership team structure lacks formal authority, then these teams are less likely to improve instruction for students (Supovitz, 2018).

### *Sustaining Professional Development Through Active Learning*

Districts can deliver both the policy environment and the resources schools need to ensure that their teachers can analyze and discuss student work using rich assessments

within their learning communities. However, whether and how these professional learning opportunities are effectively adopted depends on the support and mentorship that school administration and leadership teams provide to teachers. Districts should know if school-based professional learning opportunities have been designed to allow teachers to use the instructional and assessment strategies and knowledge discussed in this section to engage in important sense-making and set goals for their learning (Coburn et al., 2013a). If these key features are missing, district administrators can partner with school leaders to strengthen their capacity to build such practices into professional learning. School leaders who shape professional learning within the context of the improvement sciences (e.g., Plan-Do-Study-Act or coaching cycles) are likely to emphasize the importance of active learning, or having teachers reflect on and refine their enactment efforts continuously (Darling-Hammond et al., 2017; Hanno, 2022; Penuel et al., 2020a). This active learning approach serves as another important characteristic of a *learning systems* district or school, where the professional learning work becomes highly adaptative, dynamic, and responsive to the varying and changing needs of students.

The vignette about Prairie Heights Middle School that opens this section describes active learning overseen by school leaders and supported by district administrators. At that school, educators collaborate in disciplinary teams to design strategies they can apply in their classrooms. Then, the teachers receive input from their mentors to help transform their teaching practice. In each subsequent meeting with their disciplinary teams and mentors, Prairie Heights Middle School teachers engage in reflection and inquiry as they consider how to refine their teaching strategies, based on their analysis of student work and their interactions with students. Another facet of active learning that is, as Darling-Hammond (2020) notes, crucial for teachers who aim to understand the development and learning of students, is to reflect on how students' personal lives and circumstances interact with and can directly influence the ways in which students learn. Paying close attention to how students' personal lives and circumstances interact with learning can provide critical information about other types of supports (e.g., social-emotional learning supports) that should be built into classroom activities to better attend to the learning needs of students (Darling-Hammond, 2020).

Active learning serves as an important characteristic for effective professional development because teachers can directly relate to, apply, and continuously refine what they learn to activities that take place in their classrooms (Darling-Hammond, 2020; Hanno, 2020; Penuel et al., 2020a). As they continually reflect upon and evaluate their enacted practices within an active learning professional context, an active learning approach provides teachers with the flexibility to embark on a "change sequence" or the ability to continuously adapt and rework planned instruction, which serves as an important marker to gauge whether they are engaging in equitable instruction (Clarke & Hollingsworth, 2002). Equitable instruction, in this case, means the application and modification of strategies that teachers continuously adapt and revise to optimize student learning.

This type of active learning professional development approach also encourages school leadership teams to consider what additional resources are needed to support the coaching cycles established by schools so that, following each cycle, teachers can see the direct benefits and applications of their new knowledge and skills. When active learning is used in coordination with coaching and mentoring cycles, the activities taken

up in these learning communities maintain coherence—teachers revisit topics and build on the knowledge they have gained in each subsequent coaching cycle. This, in turn, helps ensure that what teachers learn becomes integrated into their regular teaching routines, rather than fading over time, as has been documented in several professional development studies (Boston & Smith, 2009; Hanno, 2022).

## Grading

In this section, we review the third critical area of *instructional infrastructure* that is needed to support instruction and assessment: grading. Although the signals associated with grading typically accompany ongoing instruction and assessment, grading practices will need to be examined to ensure that they support rather than work against ambitious teaching. We begin with another vignette to illustrate how one district engaged parents, students, and educators to transform a grading system that can better support student learning. This vignette (see Box 6-3) was assembled from conversations with Lori Cooper from Fountain Fort Carson School District who provided information about this ongoing work to shift grading practices.

We discuss grading as an important component of *instructional infrastructure* as it is tied to a set of district- and school-defined policies and practices that lie outside those governing the scoring of classroom-based and distal assessments. Feldman (2023) states that grades serve as a powerful means for communicating student progress and also represent one of the few areas of teaching that falls directly within an individual teacher’s control. As a result, grading practices can vary widely among teachers, even within the same school, and often vary across schools. In some situations, grading practices can create environments that decrease student motivation to learn. Within the framework of an *instructional infrastructure*, pursuing equitable grading policies and guidance for implementation provides an opportunity for districts and schools to support and align grading practices with the values that undergird how to provide feedback to students in a learning-centered vision.

Changing the way grading takes place in the classroom requires districts to provide leaders with guidance and flexibility so that they can determine the best way to shift grading practices among their teachers. Additionally, given both the role of grades as a monitoring device for many parents and how school structures have been set up in support of this purpose, challenges and potential pushback from some parents may be unavoidable—which means that clear communication is necessary, as is inviting parents into the conversation to provide them with the reasoning behind shifts in grading practice.

Many scholars have pointed to the inherent challenges in shifting traditional grading practices at schools, especially as these challenges relate to teacher attitudes and perceptions about grading (Guskey, 2021; McMillan, 2001; McMillan et al., 2002). The FFC vignette that begins this section shows a district that is intentional about bringing teachers into conversations about grading practices. By inviting teachers to discuss their thoughts and criticisms about existing practices, the district provided them with an opportunity to reflect and begin correcting some of the more problematic practices deployed in their classrooms. This example highlights one way that districts and

### **BOX 6-3**

#### **Fountain Fort Carson School District: Shifting Grading Practices**

Fountain Fort Carson School District (FFC), located in Colorado Springs, Colorado, recently adopted standards-based grading practices. As the district began its transition to a more equitable grading practice, it determined that a philosophical shift around grading would need to occur in order to provide a rationale for the practices that would follow. The district revisited existing and common grading practices such as: using averages to determine final grades, including student behavior as a part of grades, and using grades as punishment. One of the first steps it took was to generate a buy-in among high school teachers for changing grading practices. Thus, FFC encouraged teachers to submit lists that described existing problematic grading practices at the district's two high schools. After a review of the input provided, school leadership and staff identified the top ten practices teachers considered to be exacerbating the inequitable evaluations of students. These practices included providing extra credit, using grades as a punitive or disciplinary tool, assigning zero grades to missing or incomplete homework, and factoring attendance into grades.

The next step the district took, in partnership with the two high schools, was to bring parents into the conversation. To gather input directly from parents about the educational experiences offered to their children, the high schools established a structure they refer to as Learning Walks. Learning Walks are an opportunity for parents to tour the school with the goal of "lifting the curtain" to showcase instructional practices and to hear from students about their learning experiences. In a panel with students during one Learning Walk, parents engaged with and heard directly from students about how the newly instituted grading practices were impacting their learning. This, in turn, helped bolster parent support for these practices.

Before instituting the new grading practices, school leaders, with district support, determined that the practices would not be implemented as a top-down mandate. School leadership made this decision because they wanted to enact sustainable and incremental changes that would intrinsically motivate their teachers to adopt the new practices. To this end, teachers were encouraged to try out the new practices, to learn by doing, and to ask students for their feedback along the way. As a result, the high schools have increased equitable grading practices in their district.

Presently, an increased number of teachers across the district use a decaying average grading system that emphasizes a student's current performance (or the most recent evidence of this performance) to determine their final grade. Homework and quizzes are treated as (ungraded) practice that provides students with opportunities to receive feedback and correct their errors and misunderstandings. Additionally, teachers no longer assign zeros for missing work. The district implemented a 0–4 scale with clear proficiency criteria for each level, and teachers grade students exclusively on their mastery of content while providing students with separate opportunities to self-assess on essential skills such as collaboration, critical thinking, and communication. As the principal of one FFC high school stated: students have voiced their support for this system in that they now experience more hope in their learning compared to in the past. One district administrator recalled a conversation with a student who said, "This is the first time I've ever had hope in my math class." The administrator explained that the student now felt motivated because they had "never passed with more than a D."

schools can begin shifting teacher mindsets about grading. Educational systems that seek to build deeper learning opportunities for all students will need to consider how to best evaluate student knowledge and competencies to meet learning-centered goals. While there are many ways to evaluate student performance, each method carries its own set of tradeoffs—students will benefit from clearly defined success criteria that also help to maintain fairness and accuracy in the evaluation of all students (Berns, 2015; Nieto, 2013).

### *Shifting Grading Practices to Support Ambitious Teaching*

Grading practices that support a vision of ambitious teaching and learning communicate a clear understanding of the skills and knowledge being assessed at any given time. One model for establishing clearer grading descriptors and criteria, which is employed by many districts and schools, is standards-based—or competency-based—grading. This approach evaluates student mastery of skills or knowledge relative to specific competencies or state standards (Brookhart, 2013a, 2013b). Researchers view this approach as a relatively fair and accurate means for grading student performance because the rubrics used are intended to provide an abundance of relevant and meaningful information that students can use to improve (Feldman, 2023; Lewis, 2020; Muñoz & Guskey, 2015). This focus on mastery is also intended to emphasize deep content learning, particularly when the learning targets and success criteria are clearly defined (Lewis, 2020).

Feldman (2023) argues that standards-based or competency-based grading approaches should provide all students with grades that are fair and meaningful, regardless of any one student’s personal context or learning needs. Feldman (2023) advises that “equitable grading is accurate, bias-resistant, and motivational” (p. 77). These pillars highlight an important guide for grading policies implemented in schools: grading practices should be an accurate reflection of what students know, and should not factor in criteria such as attendance, behavior, or completed homework. In practice, this would mean that teachers ensure that grades reflect criteria that are fair and transparent to all students and are weighted to reflect a student’s most recent performance as opposed to averaging their performance over time (Feldman, 2023). Within this model, teachers provide students with rubrics that contain clear descriptors, or they co-create rubric criteria with their students, allowing students to demonstrate their knowledge through various formats and modalities over time, and provide opportunities for revisions and improvement (Brookhart, 2013b; Feldman, 2023; Nieto, 2013).

In addition to shifting toward clear, qualitatively established, criteria-based grading that has been developed in consultation with teachers, district and school leaders may need to revisit established policies for entering grades into learning management or information systems. In the specific case of FFC, teachers recognized that they were grading students on assignments that were intended to encourage risk-taking and learning from mistakes. Grading those assignments undermined this intention since grading elevated the stakes associated with submitted work. Revisiting the types of assignments and products graded and entered into the learning management system is critical for ensuring that the appropriate form of assessment aligns with and supports the underlying learning-centered values of ambitious teaching.



### *Transparency for Grading Practices and Policy Shifts*

Along with establishing guidance on the use of equitable grading practices within their schools, a *learning system* district or school should engage the school community in policy decisions that would shift grading practices. In the same way that teachers may be resistant to enacting new grading approaches, students and families need adequate time to both fully understand and adjust to these new practices (Hany et al., 2016; Townsley, 2019). Schools that provide students and parents with opportunities to engage with these changes can increase school and family collaboration, engagement, and trust in their implementation (Henderson & Mapp, 2002; Stosich & Bae, 2018). The FFC vignette that begins this section shows one way to do this: the Learning Walks provided families with an opportunity to gather information about changes to grading practices and to see them in action. FFC also provided families with an opportunity to hear directly from students about the positive impact of these changes. Attempts to change grading practices are often unsuccessful because district or school leaders tend to enact them before being transparent about the rationale for the changes being made (Guskey, 2021). In addition, Guskey (2021) notes, leaders may not allow for adequate discussion on how the new practices align with other parts of the educational process (i.e., curriculum, instruction, assessment). While such clarification may not avoid all potential opposition, Guskey (2021) argues that communicating and soliciting feedback from stakeholders can help leaders address the challenges that may lie ahead “with patience, purpose, and resolve” (p. 196).

In summary, fair and equitable grading policies and practices complement work in other parts of the *instructional infrastructure* that supports ambitious teaching. If grading practices are not considered when adjusting the *instructional infrastructure*, existing practices could potentially work against efforts to promote equitable assessment practices. Steps that can be taken to promote equitable grading practices include:

- Avoiding grading methods that demotivate students and work against ambitious teaching and learning practices (e.g., assigning zeros to unsubmitted work, factoring participation and attendance into grades, using 100-point grading scales, and grading non-academic or soft skills such as collaboration).
- Re-evaluating the body of evidence used for grading and ensuring that instructional tasks and assignments used for formative purposes are not factored into grades.
- Providing students with multiple opportunities to demonstrate their mastery of a skill.
- Developing clear criteria in grading rubrics and ensuring that they connect to expected learning targets and goals. Ideally, criteria should be co-constructed with students to ensure a shared understanding.

District and school leaders who provide policy guidance to teachers can help establish consistency in grading practices by creating transparency into these practices and by encouraging the creation of learning environments in which expectations of high-quality teaching and learning are present for both teachers and students (Brookhart, 2013a; Diefes-Dux, 2018; Feldman, 2023).

## TAKING A *LEARNING SYSTEMS* APPROACH TO IMPLEMENTING A LEARNING-CENTERED VISION

Enacting the practices and policies described in this chapter requires significant district investment, time, and resources. The transition, for districts, into the *learning systems* stage, focused on advancing ambitious teaching and learning, is a journey. This journey can be daunting, especially when considering the amount of coordination and support work that is required across classrooms and schools. For schools taking up this vision for the first time, the journey can also feel overwhelming, as it requires attending to many areas, including revisiting the school schedule to make it possible for teachers to engage in meaningful professional learning and collaboration.

For districts and schools moving in the *learning systems* direction, the work might begin gradually, in one disciplinary area, one or two grades, just a few classrooms, or a few schools. Pursuing incremental changes will be more manageable for all involved, and will likely help avoid “shallow” results, which are more likely if a transition is implemented at scale and all at once (Cook-Endres et al., 2014). Districts leading the work to implement infrastructure changes may also choose to implement a manageable, small pilot in a single school, or for a single grade level or content area, to learn deeply from this work. Indeed, by using effective partnership models, districts and schools will also facilitate or improve their ability to learn from and improve on this work together at a manageable scale, before considering and potentially re-evaluating steps needed to spread the work gradually and incrementally into other grades, content areas, and schools (see Chapter 8 of this volume, “Developing, Implementing, and Institutionalizing Complex Educational Innovations: Considerations for Balanced Assessment Systems”). Additionally, including the voices of community members in these decisions should also broaden ownership and sustainability in these types of reform (Arriaza, 2004; Penuel et al., 2020b).

For districts and schools wanting to move into a *learning system* phase, it is critical to consider what to prioritize or refine in the existing infrastructure to support ambitious teaching and learning. Prioritization is important because districts and schools face resource constraints that limit what work can be taken up at a given time. Addressing just one area of the *instructional infrastructure* (e.g., providing high-quality curriculum materials) requires significant investment and work. For districts seeking to build on existing infrastructure, drawing on improvement science approaches to identify priorities in consultation with school- and community-based stakeholders, including students, will be helpful (Penuel et al., 2020b).

Improvement science approaches, such as design thinking or Plan-Do-Study-Act cycles, integrated into a community-based design are particularly useful for facilitating the adoption or the improvement of reforms enacted. First, these approaches can provide an effective process for districts and schools to engage stakeholders in appraising the existing infrastructure and identifying areas to prioritize. Many schools and districts already use these types of approaches as process evaluations for improving instructional strategies, including assessment. Second, these collaborative approaches provide districts with opportunities to ensure that school-based personnel, students, and the broader community can help codesign or provide substantive input into the areas of infrastructure that require rethinking and additional investment. Lastly, these

approaches provide natural opportunities for districts and schools to experiment and learn from infrastructure adjustments and changes.

There is no “right” or specific pathway to approach the work of becoming a *learning systems* district or school, centering culturally responsive and learning-centered classroom activity systems. However, districts and schools that seek to become *learning systems* can consider engaging in some of the activities below to begin the process of identifying and determining areas to strengthen or transform:

- Develop a theory of action or change as a starting point to help define the implementation work and supports needed to build or bolster one or more components of the *instructional infrastructure*. This development work could include conducting interviews with a broad range of stakeholders to understand the existing barriers, challenges, and opportunities encountered in each infrastructure area.
- Undertake a review and /or audit of curriculum materials, inclusive of assessments, used in schools to determine the extent to which materials can support teachers in enacting ambitious teaching and learning. Results from this review can inform which materials and tasks are good candidates for adaptation and identify those that may need to be replaced.
- Establish research partnerships with organizations, including higher education institutions, to build an evaluation plan for engaging in studies that will enable districts and schools to learn from their continuous improvement work.

In reference to the third bullet, whether a district is starting this work or building on existing infrastructure, it is necessary to establish an evaluation plan for all areas impacted by the change. The plan should focus on building ownership of this work with schools and the broader community, if the efforts are to be sustained and if the district is to maximize opportunities for all students to learn (Arriaza, 2004; Coburn, 2003; Penuel et al., 2020b). This continuous learning will need to be monitored to determine if policies and resources directed at a particular area of the infrastructure are ultimately leading to expected outcomes. Districts may decide that this learning work should focus on monitoring areas such as the maintenance of access to and provision of high-quality curriculum materials and other resources. Alternatively, planned evaluation work may focus on preventing the development of separate learning tracks (e.g., removing policies and existing barriers for students to access higher-level course opportunities). It could also attend to monitoring how and whether teachers are taking up professional learning in their classrooms or prioritize the evaluation of how implemented policies and practices are expanding deeper learning opportunities for all students.

## CONCLUSION

Moving toward a learning-centered vision that supports ambitious teaching and learning will require stakeholders to take up the courageous and difficult work of disrupting and dismantling existing infrastructures that continue to perpetuate student inequities in many districts and schools across the United States. This work is especially salient now, after disruptions to schooling caused by the COVID-19 pandemic, which

exacerbated existing inequalities in educational opportunities and learning on a global scale (Barron Rodriguez et al., 2021; Nicola et al., 2020; United Nations Educational, Scientific and Cultural Organization et al., 2021). Although attempts to address these inequalities have sparked well-intentioned efforts to reignite learning at schools, some states, districts, and schools have renewed their efforts to reinvigorate infrastructures like those seen in the NCLB era. For example, there is a growing interest across states to develop through-year assessments that entail heavy investments in vendor-developed benchmark and interim tests (Marion, 2021). Marion (2021) notes that these through-year assessments are being developed to serve multiple purposes including providing predictive information about student performance on the state end of year test during the school year, as well as providing “instructionally useful” information to students and teachers. We have also observed a few districts beginning to reshape programming offered to students by increasing instructional time dedicated to tested subjects (English language arts and math) and refamiliarizing students with testing strategies to accelerate learning from learning losses attributed to the pandemic. Increases in learning time for these two tested subjects come at the expense of programming for students in the sciences, social studies, and other disciplinary areas that contribute to and enhance student well-being and learning.

State, district, and school leaders should heed the fact that scant evidence supports the idea that short-term gains achieved by enacting test-supportive teaching and learning practices will lead to deeper learning (McTighe & Gareis, 2021). Instead, evidence shows that collectively building and sustaining instructional infrastructure designed to advance ambitious teaching will, in the long run, provide students with the knowledge and skills to address urgent societal issues given the cross-cutting tensions impacting our social, racial, economic, environmental, and political climate today. This means cultivating thriving classroom activity systems that motivate students to build a repertoire of knowledge, skills, and confidence to productively engage with others as citizens, participants, and contributors to communities at the local, national, and global levels.

## REFERENCES

- Anderson, E., & Young, M. D. (2018). If they knew then what we know now, why haven't things changed? An examination of district effectiveness research. *Frontiers in Education*, 3. <https://doi.org/10.3389/educ.2018.00087>.
- Armstrong, A. L. (2021, December 1). *The representation of social groups in U.S. educational materials and why it matters*. New America. <https://www.newamerica.org/education-policy/reports/the-representation-of-social-groups-in-u-s-educational-materials-and-why-it-matter>.
- Arriaza, G. (2004). Making changes that stay made: School reform and community involvement. *The High School Journal*, 87(4), 10–24. <https://doi.org/10.1353/hsj.2004.0007>.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258–267. <https://doi.org/10.3102/0013189X07306523>.
- Aycock, B., Drew, S., Housman, G., Ivory, S., McGill, T. A. W., Rosvally, H., Servino, M., & Wampler, R. (2019). *Why do some things wash up on the beach and others don't?* Next Generation Science Storylines. <https://www.nextgenstorylines.org/why-do-some-things-wash-up-on-the-beach-and-others-dont>.
- Barron Rodriguez, M., Cobo, C., Munoz-Najar, A., & Sanchez Ciarrusta, I. (2021). *Remote learning during the global school lockdown: Multi-country lessons*. The World Bank. <http://hdl.handle.net/10986/36141>.
- Berns, R. (2015). *Child, family, school, community: Socialization and support* (10th ed.). Cengage Learning.
- Blazar, D., & Pollard, C. (2017). Does test preparation mean low-quality instruction? *Educational Researcher*, 46(8), 420–433. <https://doi.org/10.3102/0013189X17732753>.



- Boston, M. D., & Smith, M. S. (2009). Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms. *Journal for Research in Mathematics Education*, 40(2), 119–156.
- Brookhart, S. M. (2013a). Grading. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 257–271). Sage.
- Brookhart, S. M. (2013b). *How to create and use rubrics for formative assessment and grading*. Association for Supervision and Curriculum Development.
- Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q., & Luppescu, S. (2009). *Organizing schools for improvement: Lessons from Chicago*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226078014.001.0001>.
- Burkhardt, H., & Schoenfeld, A. (2019). Formative assessment in mathematics. In H. L. Andrade, R. E. Bennett, & C. J. Cizek (Eds.), *Handbook of formative assessment in the disciplines*. Routledge.
- Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, 18(8), 947–967. [https://doi.org/10.1016/S0742-051X\(02\)00053-7](https://doi.org/10.1016/S0742-051X(02)00053-7).
- Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, 32(6), 3–12. <https://doi.org/10.3102/0013189X032006003>.
- Coburn, C. E., Toure, J., & Yamashita, M. (2009). Evidence, interpretation, and persuasion: Instructional decision making at the district central office. *Teachers College Record*, 111(4), 1115–1161.
- Coburn, C. E., Mata, W. S., & Choi, L. (2013a). The embeddedness of teachers' social networks: Evidence from a study of mathematics reform. *Sociology of Education*, 86(4), 311–342. <https://doi.org/10.1177/0038040713501147>.
- Coburn, C. E., Penuel, W. R., & Geil, K. E. (2013b). *Research-practice partnerships: A strategy for leveraging research for educational improvement in school districts*. William T. Grant Foundation. <https://files.eric.ed.gov/fulltext/ED568396.pdf>.
- Cohen, D. K., Peurach, D. J., Glazer, J. L., Gates, K. E., & Goldin, S. (Eds.). (2013). *Improvement by design: The promise of better schools*. The University of Chicago Press.
- Conley, D. (2015). A new era for educational assessment. *Education Policy Analysis Archives*, 23, 8. <https://doi.org/10.14507/epaa.v23.1983>.
- Cook-Endres, T., Taylor, A., & Penuel, B. (2014). *Using curriculum adaptation as a strategy to help teachers learn about NGSS and developing aligned instructional materials*. Brief 5. STEM teaching tools. <https://stemteachingtools.org/brief/5>.
- Crocco, M. S., & Costigan, A. T. (2007). The narrowing of curriculum and pedagogy in the age of accountability urban educators speak out. *Urban Education*, 42(6), 512–535. <https://doi.org/10.1177/0042085907304964>.
- Darling-Hammond, L. (2004). Inequality and the right to learn: Access to qualified teachers in California's public schools. *Teachers College Record*, 106(10), 1936–1966. <https://doi.org/10.1111/j.1467-9620.2004.00422.x>.
- Darling-Hammond, L. (2020). Culture, learning, and policy. In N. S. Nasir, C. D. Lee, R. Pea, & M. McKinney de Royston (Eds.), *Handbook of the cultural foundations of learning* (pp. 404–426). Routledge. <https://doi.org/10.4324/9780203774977-28>.
- Darling-Hammond, L., & Adamson, F. (2014). *Beyond the bubble test: How performance assessments support 21st century learning*. Jossey-Bass & Pfeiffer Imprints, Wiley.
- Darling-Hammond, L., & Cook-Harvey, C. M. (2018). *Educating the whole child: Improving school climate to support student success*. Learning Policy Institute. <https://doi.org/10.54300/145.655>.
- Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. <https://doi.org/10.54300/122.311>.
- Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science*, 24(2), 97–140. <https://doi.org/10.1080/10888691.2018.1537791>.
- Diaz-Bilello, E., York, A., & Mork, K. (2022). *Learnings from a multi-site case study of former turnaround schools*. Center for Assessment, Design, Research and Evaluation, University of Colorado Boulder. <https://www.colorado.edu/cadre/2022/08/12/learnings-multi-site-case-study-former-turnaround-schools>.
- Diefes-Dux, H. A. (2018). Student self-reported use of standards-based grading resources and feedback. *European Journal of Engineering Education*, 44(6), 838–849. <https://doi.org/10.1080/03043797.2018.1483896>.



- Dresser, R. (2012). The impact of scripted literacy instruction on teachers and students. *Issues in Teacher Education*, 21(1), 71–87.
- Duncan-Owens, D. (2009). Scripted reading programs: Fishing for success. *Principal*. [https://www.naesp.org/sites/default/files/resources/2/Principal/2009/J-F\\_p26.pdf](https://www.naesp.org/sites/default/files/resources/2/Principal/2009/J-F_p26.pdf).
- Dyer, K. (2017, July 13). *Understanding formative, interim, and summative assessments and their role in student learning*. NWEA. <https://www.nwea.org/blog/2017/understanding-formative-interim-summative-assessments-role-student-learning>.
- Elmore, R. F., & Burney, D. (1997). *Investing in teacher learning: Staff development and instructional improvement in Community School District #2, New York City*. National Commission on Teaching & America's Future. <https://files.eric.ed.gov/fulltext/ED416203.pdf>.
- 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](https://www.rand.org/pubs/research_reports/RR354.html).
- Feldman, J. (2023). *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms* (2nd ed.). Corwin.
- Fullan, M., & Quinn, J. (2015). *Coherence: The right drivers in action for schools, districts, and systems*. Corwin: A Sage Company.
- Gitomer, D. H., & Iwatani, E. (2022). Two communities' views on test fairness. *Educational Assessment*, 27(2), 197–203. <https://doi.org/10.1080/10627197.2022.2087624>.
- Guskey, T. R. (2021). Learning from failures: Lessons from unsuccessful grading reform initiatives. *NASSP Bulletin*, 105(3), 192–199. <https://doi.org/10.1177/01926365211029375>.
- Hamilton, L. S., Stecher, B. M., & Yuan, K. (2008). *Standards-based reform in the United States: History, research, and future directions*. RAND Corporation. [https://www.rand.org/content/dam/rand/pubs/reprints/2009/RAND\\_RP1384.pdf](https://www.rand.org/content/dam/rand/pubs/reprints/2009/RAND_RP1384.pdf).
- Hanno, E. C. (2022). Immediate changes, trade-offs, and fade-out in high-quality teacher practices during coaching. *Educational Researcher*, 51(3), 173–185. <https://doi.org/10.3102/0013189X211062896>.
- Hany, K., Proctor, M., Wollenweber, J., & Al-Bataineh, A. (2016). Teacher perception of standards-based grading: Implication and effectiveness. *Journal of Teaching and Education*, 5(1), 17.
- Hargreaves, A. (2000). Mixed emotions: Teachers' perceptions of their interactions with students. *Teaching and Teacher Education*, 16(8), 811–826. [https://doi.org/10.1016/S0742-051X\(00\)00028-7](https://doi.org/10.1016/S0742-051X(00)00028-7).
- Heiser, P., Simidian, G., Albert, D., Garruto, J., Catucci, D., Faustino, P., McCarten, K., & Caci, K. (2015). *Anxious for success: High anxiety in New York's schools*. New York Association of School Psychologists and New York State School Boards Association. [https://www.nyssba.org/clientuploads/nyssba\\_pdf/Test\\_Anxiety\\_Report.pdf](https://www.nyssba.org/clientuploads/nyssba_pdf/Test_Anxiety_Report.pdf).
- Henderson, A. T., & Mapp, K. L. (2002). *A new wave of evidence: The impact of school, family, and community connections on student achievement*. National Center for Family and Community Connections with Schools: SEDL. <https://sedl.org/connections/resources/evidence.pdf>.
- Hofman, P., Goodwin, B., & Kahl, S. (2015). *Re-balancing assessment: Placing formative and performance assessment at the heart of learning and accountability*. McREL International. <https://files.eric.ed.gov/fulltext/ED568906.pdf>.
- Honig, M. I. (2013). *From tinkering to transformation: Strengthening school district central office performance*. American Enterprise Institute. <https://www.aei.org/research-products/report/from-tinkering-to-transformation-strengthening-school-district-central-office-performance>.
- Honig, M. I., & Coburn, C. (2008). Evidence-based decision making in school district central offices: Toward a policy and research agenda. *Educational Policy*, 22(4), 578–608. <https://doi.org/10.1177/0895904807307067>.
- Honig, M. I., & Rainey, L. R. (2015). *How school districts can support deeper learning: The need for performance alignment*. Jobs for the Future. <https://files.eric.ed.gov/fulltext/ED560756.pdf>.
- Honig, M. I., & Venkateswaran, N. (2012). School–central office relationships in evidence use: Understanding evidence use as a systems problem. *American Journal of Education*, 118(2), 199–222. <https://doi.org/10.1086/663282>.

- Hopkins, M., & Spillane, J. P. (2015). Conceptualizing relations between instructional guidance infrastructure (IGI) and teachers' beliefs about mathematics instruction: Regulatory, normative, and cultural-cognitive considerations. *Journal of Educational Change*, 16(4), 421–450. <https://doi.org/10.1007/s10833-015-9257-1>.
- Hopkins, M., & Woulfin, S. L. (2015). School system (re)design: Developing educational infrastructures to support school leadership and teaching practice. *Journal of Educational Change*, 16(4), 371–377. <https://doi.org/10.1007/s10833-015-9260-6>.
- Kang, H., & Furtak, E. M. (2021). Learning theory, classroom assessment, and equity. *Educational Measurement: Issues and Practice*, 40, 73–82. <https://doi.org/10.1111/emip.12423>.
- Kang, H., Windschitl, M., Stroupe, D., & Thompson, J. (2016). Designing, launching, and implementing high quality learning opportunities for students that advance scientific thinking. *Journal of Research in Science Teaching*, 53(9), 1316–1340. <https://doi.org/10.1002/tea.21329>.
- Kaufman, J. H., Tosh, K., & Mattox, T. (2020). *Are U.S. teachers using high-quality instructional materials?* RAND Corporation. <https://doi.org/10.7249/RR2575.11-1>.
- Kraft, M. A., Simon, N. S., & Lyon, M. A. (2021). *Sustaining a sense of success: The protective role of teacher working conditions during the COVID-19 pandemic*. Annenberg Institute, Brown University. <https://doi.org/10.26300/35nj-v890>.
- Ladson-Billings, G. (2014). Culturally relevant pedagogy 2.0: a.k.a. the remix. *Harvard Educational Review*, 84(1), 74–84. <https://doi.org/10.17763/haer.84.1.p2rj131485484751>.
- Latham, B. (2018, March 22). Looking for lasting change? Start talking to other departments. EdSurge. <https://www.edsurge.com/news/2018-03-22-looking-for-lasting-change-start-talking-to-other-departments>.
- Leithwood, K., Sun, J., & McCullough, C. (2019). How school districts influence student achievement. *Journal of Educational Administration*, 57(5), 519–539. <https://doi.org/10.1108/JEA-09-2018-0175>.
- Lewis, D. (2020). Gender effects on re-assessment attempts in a standards-based grading implementation. *PRIMUS*, 30(5), 539–551. <https://doi.org/10.1080/10511970.2019.1616636>.
- Linn, R. L., & Burton, E. (1994). Performance-based assessment: Implications of task specificity. *Educational Measurement: Issues and Practice*, 13(1), 5–8. <https://doi.org/10.1111/j.1745-3992.1994.tb00778.x>.
- Loewus, L. (2017, December 19). *Majority of teachers say reforms have been “too much.”* EducationWeek. <https://www.edweek.org/teaching-learning/majority-of-teachers-say-reforms-have-been-too-much/2017/12>.
- Mac Iver, M. A., & Farley, E. (2003). *Bringing the district back in: The role of the central office in improving instruction and student achievement*. Johns Hopkins University and Howard University. <https://files.eric.ed.gov/fulltext/ED481804.pdf>.
- Marion, S. (2021, April 29). *It might just be a pile of bricks! The challenges of creating balanced assessment systems*. Nation Center for the Improvement of Educational Assessment. <https://www.nciea.org/blog/it-might-just-be-a-pile-of-bricks>.
- Marion, S. (2021, September 29). *Trying to serve multiple uses with through year assessments. Assumptions, claims, and the evidence needed to support them*. National Center for the Improvement of Educational Assessment. <https://www.nciea.org/blog/trying-to-serve-multiple-uses-with-through-year-assessments>.
- Marsh, J. A. (2002). How districts relate to states, schools, and communities: A review of emerging literature. In A. M. Hightower, M. S. Knapp, J. A. Marsh, & M. W. McLaughlin (Eds.), *School districts and instructional renewal* (pp. 25–40). Teachers College Press.
- Massell, D., & Goertz, M. E. (2002). District strategies for building instructional capacity. In A. M. Hightower, M. S. Knapp, J. A. Marsh, & M. W. McLaughlin (Eds.), *School districts and instructional renewal* (pp. 43–60). Teachers College Press.
- McLaughlin, M., & Talbert, J. (2002). Reforming districts. In A. M. Hightower, M. S. Knapp, J. A. Marsh, & M. W. McLaughlin (Eds.), *School districts and instructional renewal* (pp. 173–192). Teachers College Press.
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20(1), 20–32. <https://doi.org/10.1111/j.1745-3992.2001.tb00055.x>.
- McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, 95(4), 203–213. <https://doi.org/10.1080/00220670209596593>.
- McTighe, J., & Gareis, C. (2021, June 10). *Assessing deeper learning after a year of change*. ASCD. <https://www.ascd.org/el/articles/assessing-deeper-learning-after-a-year-of-change>.

- Mehta, J., & Fine, S. (2015). Bringing values back in: How purposes shape practices in coherent school designs. *Journal of Educational Change*, 16, 483–510. <https://doi.org/10.1007/s10833-015-9263-3>.
- Meyers, C. V. (2020). An urban district's struggle to preserve school turnaround change. *Urban Education*. <https://doi.org/10.1177/0042085920966031>.
- Milosovic, S. (2007). Building a case against scripted reading programs. *Education Digest: Essential Readings Condensed for Quick Review*, 73(1), 27–30.
- Muñoz, M. A., & Guskey, T. R. (2015). Standards-based grading and reporting will improve education. *Phi Delta Kappan*, 96(7), 64–68. <https://doi.org/10.1177/0031721715579043>.
- Next Generation Science Storylines. (n.d.). *What are storylines?* <https://www.nextgenstorylines.org/what-are-storylines>.
- Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, M., & Agha, R. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International Journal of Surgery*, 78, 185–193. <https://doi.org/10.1016/j.ijisu.2020.04.018>.
- Nieto, S. (2013). *Finding joy in teaching students of diverse backgrounds: Culturally responsive and socially just practices in U.S. classrooms*. Heinemann.
- No Child Left Behind Act of 2001, 20 U.S.C. 6301. (2001). <https://www.congress.gov/bill/107th-congress/house-bill/1/text>.
- Olson, L., & Jerald, C. (2020). *The big test: The future of statewide standardized assessments*. FutureEd. <https://www.future-ed.org/the-big-test-the-future-of-statewide-standardized-testing>.
- Penuel, W. R. (2019). Infrastructuring as a practice of design-based research for supporting and studying equitable implementation and sustainability of innovations. *Journal of the Learning Sciences*, 28(4–5), 659–677. <https://doi.org/10.1080/10508406.2018.1552151>.
- Penuel, W. R., Briggs, D. C., Davidson, K. L., Herlihy, C., Sherer, D., Hill, H. C., Farrell, C., & Allen, A.-R. (2017). How school and district leaders access, perceive, and use research. *AERA Open*, 3(2). <https://doi.org/10.1177/2332858417705370>.
- Penuel, W. R., Bell, P., & Neill, T. (2020a). Creating a system of professional learning that meets teachers' needs. *Phi Delta Kappan*, 101(8), 37–41. <https://doi.org/10.1177/0031721720923520>.
- Penuel, W. R., Riedy, R., Barber, M. S., Peurach, D. J., LeBouef, W. A., & Clark, T. (2020b). Principles of collaborative education research with stakeholders: Toward requirements for a new research and development infrastructure. *Review of Educational Research*, 90(5), 627–674. <https://doi.org/10.3102/0034654320938126>.
- Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, 28(3), 5–13. <https://doi.org/10.1111/j.1745-3992.2009.00149.x>.
- Polikoff, M. (2021). *Beyond standards: The fragmentation of education governance and the promise of curriculum reform*. Harvard Education Press.
- Randall, J., Slomp, D., Poe, M., & Oliveri, M. E. (2022). Disrupting white supremacy in assessment: Toward a justice-oriented, antiracist validity framework. *Educational Assessment*, 27(2), 170–178. <https://doi.org/10.1080/10627197.2022.2042682>.
- Reiser, B. J., Novak, M., McGill, T. A. W., & Penuel, W. R. (2021). Storyline units: An instructional model to support coherence from the students' perspective. *Journal of Science Teacher Education*, 32(7), 805–829. <https://doi.org/10.1080/1046560X.2021.1884784>.
- Ruiz-Primo, M. A., Li, M., Wills, K., Giamellaro, M., Lan, M.-C., Mason, H., & Sands, D. (2012). Developing and evaluating instructionally sensitive assessments in science. *Journal of Research in Science Teaching*, 49(6), 691–712. <https://doi.org/10.1002/tea.21030>.
- Schoenfeld, A. H. (2013). Classroom observations in theory and practice. *ZDM*, 45, 607–621. <https://doi.org/10.1007/s11858-012-0483-1>.
- Schoenfeld, A. H. (2017). Uses of video in understanding and improving mathematical thinking and teaching. *Journal of Mathematics Teacher Education*, 20, 415–432. <https://doi.org/10.1007/s10857-017-9381-3>.
- Shepard, L. A. (2017). Formative assessment: Caveat emptor. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 279–303). Routledge. <https://doi.org/10.4324/9781315086545>.
- Shepard, L. A. (2019). Classroom assessment to support teaching and learning. *ANNALS of the American Academy of Political and Social Science*, 683(1), 183–200. <https://doi.org/10.1177/0002716219843818>.

- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, 37(1), 21–34. <https://doi.org/10.1111/emip.12189>.
- Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement*. MDRC. [https://www.mdrc.org/sites/default/files/foundations\\_for\\_success\\_summary.pdf](https://www.mdrc.org/sites/default/files/foundations_for_success_summary.pdf).
- Spillane, J. P., Healey, K., Parise, L. M., & Kenney, A. (2011). A distributed perspective on learning leadership. In J. Robertson & H. Timperley (Eds.), *Leadership and learning* (pp. 159–171). SAGE Publications. <https://doi.org/10.4135/9781446288931.n12>.
- Stosich, E. L., & Bae, S. (2018). Engaging diverse stakeholders to strengthen policy. *Phi Delta Kappan*, 99(8), 8–12. <https://doi.org/10.1177/0031721718775670>.
- Suárez, E., Bell, P., McCulloch, A., & Starr, M. (2020). *Why you should stop pre-teaching science vocabulary and focus on students developing conceptual meaning first*. STEM Teaching Tools. <https://stemteachingtools.org/brief/66>.
- Supovitz, J. A. (2018). Teacher leaders' work with peers in a quasi-formal teacher leadership model. *School Leadership & Management*, 38(1), 53–79. <https://doi.org/10.1080/13632434.2017.1389718>.
- Teoh, M., Coggins, C., Guan, C., & Hiler, T. (2014). *The student & the stopwatch: How much time do American students spend on testing?* Teach Plus. [https://teachplus.org/wp-content/uploads/files/publication/pdf/the\\_student\\_and\\_the\\_stopwatch.pdf](https://teachplus.org/wp-content/uploads/files/publication/pdf/the_student_and_the_stopwatch.pdf).
- Thompson, J., Mawyer, K., Johnson, H., Scipio, D., & Luehmann, A. (2021). *C<sup>2</sup>AST (Critical and Cultural Approaches to Ambitious Science Teaching): From responsive teaching toward developing culturally and linguistically sustaining science teaching practices*. National Science Teaching Association. <https://www.nsta.org/science-teacher/science-teacher-septemberoctober-2021/c2ast-critical-and-cultural-approaches>.
- Togneri, W., & Anderson, S. E. (2003). *Beyond islands of excellence: What districts can do to improve instruction and achievement in all schools—A leadership brief*. Learning First Alliance. <https://files.eric.ed.gov/fulltext/ED475875.pdf>.
- Townsley, M. (2019). Considering standards-based grading: Challenges for secondary school leaders. *Journal of School Administration Research and Development*, 4(1), 35–38.
- United Nations Educational, Scientific and Cultural Organization, UNESCO Institute for Statistics, United Nations Children's Fund, UNICEF Office of Research, The World Bank, & Organisation for Economic Co-operation and Development. (2021). *What's next? Lessons on education recovery: Findings from a survey of ministries of education amid the COVID-19 pandemic*. <https://www.oecd.org/education/what-s-next-lessons-on-education-recovery-697bc36e-en.htm>.
- Wang, E. L., Tuma, A. P., Doan, S., Henry, D., Lawrence, R. A., Woo, A., & Kaufman, J. H. (2021). *Teachers' perceptions of what makes instructional materials engaging, appropriately challenging, and usable: A survey and interview study*. RAND Corporation. <https://doi.org/10.7249/RRA134-2>.
- Watson, C. (2014). Effective professional learning communities? The possibilities for teachers as agents of change in schools. *British Educational Research Journal*, 40(1), 18–29. <https://doi.org/10.1002/berj.3025>.
- Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd ed.). Association for Supervision and Curriculum Development.
- Willis, L., Badrinarayan, A., & Martinez, M. (2022). *Quality criteria for systems of performance assessment for school, district, and network leaders*. Learning Policy Institute. <https://doi.org/10.54300/439.730>.
- Windschitl, M., Thompson, J., Braaten, M., & Stroupe, D. (2012). Proposing a core set of instructional practices and tools for teachers of science. *Science Education*, 96(5), 878–903. <https://doi.org/10.1002/sce.21027>.
- Windschitl, M., Thompson, J., & Braaten, M. (2018). *Ambitious science teaching*. Harvard Education Press.



# State Practices and Balanced Assessment Systems

Carla Evans, *National Center for the Improvement of Educational Assessment*

Erika Landl, *National Center for the Improvement of Educational Assessment*

## CONTENTS

INTRODUCTION .....	201
BACKGROUND .....	204
High-Stakes Federal Accountability .....	204
Federal Requirements .....	205
Outsized Role of State Assessments in Shaping Local Curriculum, Instruction, and Assessment .....	205
STATE ACTIONS CONTRIBUTING TO IMBALANCED ASSESSMENT SYSTEMS .....	206
Layering on Additional High-Stakes Accountability Decisions Based on State Assessments .....	206
Unclear Communications About How State Assessments Will Be Used to Inform Decisions or Actions .....	206
Failure to Envision State Assessments Within the Broader System of Assessments.....	207
BALANCED ASSESSMENT SYSTEMS: WHO CONTROLS WHAT?.....	208
Local Control .....	208
State Control.....	209
STATE ACTIONS TO SUPPORT BALANCED ASSESSMENT SYSTEMS.....	210
State Action 1: Set a Clear, Compelling, and Coherent Theory of Action for Balanced Assessment Systems.....	211
State Action 2: Clearly Communicate the Intended Role of the State Summative Assessment and Other State-Provided Resources Within Balanced Assessment Systems.....	215
State Action 3: Proactively Design State Content Standards, Curriculum Frameworks, and State Assessments to Promote Coherence.....	216
State Action 4: Mitigate Misuse of the State Test Through Clear Reporting and Guidance .....	221
State Action 5: Provide Tools, Resources, and Support to LEAs.....	222
State Action 6: Engage Educator Preparation Programs .....	231
CONCLUSION .....	232
REFERENCES .....	233



## INTRODUCTION

In this chapter, we argue that the primary role of states in promoting balanced assessment systems should be to create and support the right structures and conditions for district and school leaders and classroom educators to effectively improve student learning. We acknowledge that states cannot design or implement balanced assessment systems on their own because they have limited control over the elements that comprise or influence local decisions. Most decisions that impact the design and implementation of local assessment systems are made at the district-, school-, and classroom-level. This is not to imply or suggest that states do not have a critical role to play in supporting more balanced assessment systems but simply acknowledges that states serve a supporting role, which represents a difference in action—not importance. The purpose of this chapter is to discuss how states can have a positive and appropriate influence on the portfolio of assessments used by educators to support ambitious teaching and equitable assessment practices.<sup>1</sup>

Several factors influence a state education agency's (SEA's) impact on local assessment policies and practices.<sup>2</sup> While some of these factors are out of a SEA's control (e.g., federal mandates), others represent a state's unique political landscape and perceived role in supporting student outcomes, both of which can influence the resources and autonomy afforded to districts and schools in making decisions about curriculum, assessment, and instruction. For example, while all states have federally required summative assessment programs,<sup>3</sup> some SEAs also provide districts and schools with optional interim assessments and associated resources, progress monitoring tools, and item banks. Similarly, local education agencies (LEAs) vary in the degree to which they value, trust, and use results from the state summative assessment program and/or other state-provided tools to improve local practices. In some contexts, the SEA is seen by the LEA as a crucial partner; in other contexts, the SEA fulfills only a monitoring function—and many state and local partnerships fall somewhere in between these two extremes.

Given the marginal influence that SEAs have on local assessment decisions, an SEA's effectiveness in promoting balanced assessment systems rests on its ability to create and promote structures, policies, and resources (e.g., tools, guidance) that (a) foster trust between state and local entities as well as reciprocal accountability; (b) incentivize practices that prioritize students' unique learning needs and academic outcomes; (c) signal what is important to teach and for students to learn; (d) promote fair, appropriate, inclusive, and equitable assessment practices; and (e) discourage and

---

<sup>1</sup> Ambitious teaching “centers on each student’s engagement and participation; it requires paying explicit attention to who students are as they enter the classroom, including their prior learning experiences (inside and outside formal educational settings), their family- and community-based funds of knowledge, and their races, ethnicities, gender identities, social classes, and other aspects that influence their identities as learners.... Equitable assessment is embedded in and enables ambitious teaching” (Shepard, 2021). See Shepard (2021) for more details.

<sup>2</sup> Throughout this chapter, we distinguish the state education agency (SEA) from other state governmental bodies (e.g., state legislature, governor’s office, the state board of education) that can influence or constrain decisions about the state assessment program and how it is viewed or used. In this chapter, the term “state” with no modifier refers to both the SEA and state governmental bodies.

<sup>3</sup> Throughout this chapter, we use the following terms interchangeably: federally required/mandated state summative testing program, state summative assessment program, state-required annual achievement tests, state tests, state testing programs, and state summative tests.

mitigate assessment practices that perpetuate systemic inequities and/or work against efforts to create rich learning environments (Chappuis et al., 2016; Conley & Darling-Hammond, 2013; National Research Council, 2010; Stiggins, 2006, 2008, 2017).

The argument laid out in this chapter rests on two overarching assumptions. The first assumption is that a state's influence on local assessment practices can and should extend beyond the state's summative assessment program. The second is that balance is best conceptualized as existing along a continuum. Each of these assumptions is discussed below.

Measures of student participation and performance on federally mandated state summative assessments are included in school accountability determinations. Consequently, factors that influence student performance on the state summative assessment are likely top of mind when school and district leaders make decisions about the materials used to drive and evaluate teaching and learning. This signaling function is beneficial to the extent that the state summative assessment clarifies the expectations underlying the state content standards, does not work against deeper learning practices, demonstrates high-quality item and test development, and demonstrates implementation practices that support appropriate test use and interpretation (e.g., attainable expectations for performance on the test have been established). However, given its influential role, the state summative assessment program can also have an outsized negative influence if not situated as one element of a broader system of assessments.

Within a school year, students participate in a broad range of assessments directed or mandated by actors at different levels of the educational system (e.g., district, school, and classroom) for different purposes (e.g., screening, instruction, evaluation). To provide information to stakeholders that will improve their decision making and positively impact teaching and learning, the assessments must work together to provide a useful and coherent profile of information about student achievement (i.e., learning strengths and needs, student performance and growth) (Marion, 2019b). Achieving this goal is no easy task—it requires the coordinated planning and engagement of multiple stakeholders and a basic understanding, at the very least, of the fundamentals of assessment design. If states do not participate in efforts to improve local assessment practices “there is a greater likelihood that assessment systems will remain incomplete or incoherent” (Gong, 2010).

A local assessment system includes all the assessments administered to students in a year in a district or school, including state-required annual achievement tests, school- and/or district-required assessments, and classroom assessments. Therefore, SEAs cannot dictate the design of local assessment systems but can provide resources and guidance that (a) clarify the intended role of the state summative assessment program and any other state-provided tools, (b) improve assessment literacy, and (c) advocate for policies, opportunities, or incentives that will foster improved local assessment practices. In some cases, SEAs will have a greater influence on local assessment practices, such as when districts and schools are identified for state support. States can also proactively address issues within their control that are likely to undermine the balance of local assessment systems, as discussed in more detail later in this chapter.

To this chapter's second assumption, balance is not dichotomous—it is a matter of degree. An assessment system cannot be unbalanced one day and balanced the next. Balance exists on a continuum, which reflects the extent to which desired

characteristics—coherence, continuity, comprehensiveness, efficiency, and utility—are represented in the set of assessments under consideration (in this volume, see Chapter 1, “Reimagining Balanced Assessment Systems: An Introduction,” and Chapter 2, “The Struggle to Implement Balanced Assessment Systems: Explanations and Opportunities”). In addition, since assessments are selected or determined by a diverse set of stakeholders, balance may be differentially represented at various levels of the educational system. For example, after conducting a local assessment system audit, a district may discontinue poorly aligned assessments and/or assessments that are not perceived as providing timely, useful information. In doing so, the district will increase the overall balance of its assessment system by improving efficiency and utility, while confirming that the remaining assessments are coherent with each other and the district-defined curriculum (e.g., reflect a common approach to learning). This does not mean, however, that the district’s assessment system will meet the needs of all stakeholders or reflect a clear, consistent message about where students need support when paired with classroom assessment information. In fact, the district’s attempt to improve efficiency may be perceived by some as negatively impacting the comprehensiveness of the system (i.e., the range of information provided to inform decision making). A separate analysis that considers school-level assessment practices in combination with the materials and assessments enacted at the classroom level may be necessary to evaluate the degree to which balance is reflected within and across these levels and how it can be improved. Consequently, improved balance at one level of the system—whether it be state, district, or school—is not sufficient to ensure that the overall system will have a positive impact on teaching and learning.

Additionally, SEA leadership cannot articulate what a district or school leader or classroom educator needs at the level of granularity necessary to design local practices around curriculum, instruction, and assessment. However, SEAs can think about the role they believe the state summative assessment—and other state-developed and -provided tools—can or should play. The SEA can then ensure that communication and resource efforts are aligned with those beliefs to appropriately inform local assessment decisions and practices.

The efforts necessary to promote balance depend on the range of information that stakeholders need to support student learning. These needs can change over time in predictable and unpredictable ways. Therefore, in the same way assessment validation requires ongoing evidence collection to support score interpretation and use, maintaining balance should be perceived as an ongoing process of adjustment rather than an attainable end state.

The principal audience for this chapter is SEA personnel who are tasked with designing and implementing the state’s vision for education through multiple means, including the state’s assessment program. A secondary audience is state legislatures, state boards of education, and state chiefs, who have significant control over education policy—including policies that can afford or constrain state assessment and accountability decisions made by SEAs and LEAs. Another key audience is test vendors, who are partially responsible for operationalizing a state’s vision for its assessment program or system.

This chapter begins by situating a state’s role in designing and implementing balanced assessment systems within a larger sociopolitical context. Specifically, we

consider how federal accountability and peer review requirements influence state assessment decisions and exert pressure on districts and schools that can trickle down to the classroom.

Subsequently, we discuss how these contextual factors result in state actions that impact the balance of assessment systems. In particular, we focus on the outsized and often unclear way that districts, schools, and classrooms use state summative assessment results to inform decisions aimed at improving student learning. We also note the lack of systems thinking demonstrated by SEAs with respect to supporting the design and implementation of high-quality assessment practices due to concerns over local control.

Next, we compare what is under local versus state control regarding the design and implementation of balanced assessment systems. This section serves to ensure that the recommendations that follow are appropriately aligned to the decisions state education agencies, state boards of education, state legislatures, and state chiefs are tasked with and can reasonably change.

Finally, the remainder of this chapter focuses on high-leverage actions states can take to promote the design and implementation of more balanced assessment systems within and across levels of the educational system.

## **BACKGROUND**

### **High-Stakes Federal Accountability**

Accountability uses of state assessment results (e.g., school accountability ratings or designations) can work against state and local efforts to develop balanced assessment systems in obvious and hidden ways (Fuhrman & Elmore, 2004). Specifically, using accountability as a driver for whole-school reform can impede ambitious teaching and equitable assessment practices because the pressure to raise student test scores will take precedence over student learning. Elevating the importance of test scores can result in the proliferation of commercial interim assessments to predict performance and monitor progress (Marion et al., 2019), teaching the test (Supovitz, 2009), narrowing of the curriculum (Au, 2007), educational triage (Booher-Jennings, 2005; Diamond & Spillane, 2004), and other pernicious practices and effects (Firestone et al., 2000). Consequently, state efforts to positively impact local assessment practices will not be successful without thoughtful reform in how federal accountability is enacted.

However, accountability is not always negative. For example, adding state summative tests in social studies can help balance the incentive to focus solely on English language arts, mathematics, and science to the exclusion of other core subject areas. Additionally, accountability regulations that require school-level academic achievement to be disaggregated by student group can highlight the achievement gaps of marginalized communities and under-represented groups like students with disabilities and English learners. For example, as advocates from these student groups would contend: “we cannot fix what we cannot measure” and their organizations “rely on the consistent, accurate, and reliable data provided by annual statewide assessments to advocate for better lives and outcomes for our children” (The Leadership Conference on Civil and Human Rights, 2015).

## Federal Requirements

Federal regulations directing the design, administration, reporting, and peer review<sup>4</sup> of state assessments create conditions that influence the role the state summative assessment can and should play in districts and schools. For example, federal requirements dictate annual state testing for students in certain grades and subject areas. These state tests must meet certain requirements related to validity, reliability, and fairness, and must also produce individual student reports that allow stakeholders to understand and address students' specific academic needs. These requirements, which are evaluated as part of the federal peer review process, are designed to ensure that the state summative assessment provides high-quality information to inform public reporting and support school accountability.

Due to these requirements, states have specific constraints that affect how they design, deliver, score, and report their summative assessments. These constraints and their impacts on the assessments may send unintended signals to local educators about how learning is best evaluated. For example, federal peer review requires states to submit evidence that assessments have been designed to support student proficiency on the breadth and depth of grade-level academic content standards, comparable across classrooms in the state. Accommodating this requirement necessitates a content sampling design that meaningfully represents the grade-level standards and supports the development of items that can be evaluated within the context of a large-scale standardized assessment (e.g., selected or short-answer responses that can be scored accurately and consistently). These prioritizations, reflected in the design of the summative assessment, can negatively influence how and what teachers teach if not accompanied by clear communication about the role and purpose of the state summative assessment and the rationale underlying its design.

Despite the challenges that the peer review process can create for state summative assessment design, the process can also have a peripheral, positive impact on efforts to support balanced assessment systems. Currently, peer review is the only process that exists for evaluating the technical quality of state-designed assessments in a comprehensive, standardized manner. Because it serves as an independent standard for quality, peer review is a useful criterion when working with state governmental bodies that may want to implement assessment policies that could undermine technical quality and inclusivity because it can be used as an argument against such policies. Furthermore, because peer review results are used to label schools, identify them for support, and inform other state decisions (promotion, grades, teacher evaluation), ensuring a state's summative assessment demonstrates technical quality is a necessity.

## Outsized Role of State Assessments in Shaping Local Curriculum, Instruction, and Assessment

State assessments take an outsized role in local contexts when educators begin reshaping local curriculum, instruction, and assessment to mimic the format or struc-

---

<sup>4</sup> It is beyond the scope of this chapter to discuss the state assessment peer review process in detail. We refer readers to the U.S. Department of Education's guide to the peer-review process here for additional context: <https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf>.



ture of the state test—especially for state tests that are mainly selected responses. The state test is one instantiation of state content standards—and a limited one, given the design features necessary to fulfill federal accountability requirements (e.g., the standardization necessary to support comparability). Additionally, political, logistical, and practical constraints often limit what can be assessed—and at what depth—in state assessments. Therefore, while the content of the state test should be mirrored in the local curriculum and instructional program because both are built from the same set of content standards, the format and structure of classroom assessments can—and likely should—be more varied and distinct from the state tests. The inclusion of performance tasks on state tests is one way that states can use the often outsized role of the state test to signal the importance of complex demonstrations and applications of learning at the local level.

### **STATE ACTIONS CONTRIBUTING TO IMBALANCED ASSESSMENT SYSTEMS**

States contribute to imbalanced assessment systems by perpetuating or inflating structures and conditions that work against efforts to administer local assessments that complement curriculum and instruction. Imbalance can occur for many reasons, including layering on additional high-stakes accountability decisions based on state assessments, unclear communications about how state assessments will be used to inform decisions or actions, and failure to envision state assessments within the broader system of assessments.

#### **Layering on Additional High-Stakes Accountability Decisions Based on State Assessments**

SEAs must adhere to federal school accountability guidelines. However, many state governing bodies go beyond federal rules and regulations, requiring the use of state summative assessment results for additional high-stakes decisions. For example, some state legislatures require student test scores to be part of teacher evaluations, even though the state test is not designed to support inferences about teacher effectiveness (American Educational Research Association, 2015). In other states, state summative assessment results have been used to determine high school graduation, third-grade promotion based on literacy performance, and other gateway decisions. Using student test results in these ways is not required under federal law and can lead to behaviors, practices, and conditions that work against efforts to support balanced assessment systems. This is not to say that all additional state accountability leads to imbalance. It depends on who is being held accountable, for what, and the evidence or theory that supports such actions.

#### **Unclear Communications About How State Assessments Will Be Used to Inform Decisions or Actions**

Each year SEAs administer tests to all students in federally required grades and content areas. Although SEAs spend exorbitant amounts of time and money ensuring

that these assessments meet federal peer review requirements, few SEAs provide a theory of action that describes, in detail, how the state assessment program is intended to drive progress or inform decisions that positively impact school quality and student outcomes. This lack, coupled with a dearth of clear communication about the primary purpose of state testing, perpetuates misconceptions about how assessment results can and should be used (e.g., by teachers to make instructional decisions). These misconceptions, in turn, fuel concerns about over-testing and the value of the state assessment when those other desired uses, such as instructional usefulness, are not supported. Dissatisfaction with state assessments and opt-out movements are the visible signs of these fractures.

Similarly, states often create policies or initiate assessment reforms that influence the design, use, or impact of the state assessment program absent a clear theory of action that defines how the changes will lead to improved teaching and learning. This lack of a clear theory of action is reflected in how some states are considering or piloting through-year assessment designs. Dadey and Gong (2017, 2021) define a through-year assessment program as having assessments that are (1) administered in multiple distinct sessions during a school year, and (2) intended to support the production and use of a summative determination of student proficiency and one additional aim. The additional aim is often instructional utility. In essence, these through-year assessment reforms are trying to make state assessments serve multiple roles—the typical monitoring and accountability role, which is federally required, and an instructional support role. What is often left underspecified, however, is how the information supplied by the through-year state assessments (e.g., raw score, achievement level, scaled score) will foster high-quality instructional actions and practices at the local level (Dadey et al., 2023). For example, what specific action(s) does the state expect classroom educators to take with the assessment results? How does the grain size and frequency of information provided serve to support that use? What assumptions must hold for it to do so effectively?

### **Failure to Envision State Assessments Within the Broader System of Assessments**

When SEAs focus their efforts and communications solely on the state summative assessment program, they give up a powerful opportunity to help stakeholders understand and appropriately situate the state test within the broader system of assessments used to collect information about student performance over a year. Choices made at one layer of the assessment system can have a trickle-down or filter-up effect that can drive imbalance by constraining or inappropriately influencing decisions and actions. States can help facilitate systems thinking by (a) modeling this practice when communicating about the intended use of state assessments in relation to locally administered assessments and (b) providing assessment design and evaluation tools, resources, and supports that promote systems thinking. Each of these topics is discussed in more detail in the section titled “State Actions to Support Balanced Assessment Systems.”

Additionally, state laws or policies related to assessment can work against teaching and learning and signal different instructional priorities from that of the content standards. For example, laws or policies that focus on or necessitate keeping state tests short

and cheap (e.g., state-defined constraints on testing times and the federal requirement to test every student every year) could lead to decisions like the elimination of writing prompts or more complex item types. However, these types of items elicit students' knowledge and skills related to the depth of the content standards. Such policies or lack of funding could result in state tests that do not appropriately signal instructional priorities around deeper learning and work against the models of learning that the SEA is trying to promote. This disconnect also exemplifies a lack of systems thinking. The state test should reflect the content standards and how they are intended to be taught. For better or for worse, what gets tested gets taught—at least to some degree (Faxon-Mills et al., 2013).

### BALANCED ASSESSMENT SYSTEMS: WHO CONTROLS WHAT?

Before focusing on SEA actions that support balanced assessment systems, it is important to differentiate what is directly in the control of the state versus what is directly in the control of LEAs. *Education is a federal interest, a state responsibility, and a local function.* So, who has control of what aspects of the educational system when decisions need to be made about the factors (e.g., education policies, resources, and actions) known to indirectly or directly influence balanced assessment systems? It is important to clarify these roles upfront so that the recommendations in this chapter for SEA actions align with what is under state control.

#### Local Control

As shown in Table 7-1, LEAs control decisions about curriculum, instruction, and local assessments, as well as defining local assessment policies and practices (see Chapter 6 of this volume, “District and School Practices and Assessments to Support a Learning-Centered Vision”). For this reason, it has been stated that the primary locus of control for the design and implementation of balanced assessment systems lies with

**TABLE 7-1** Areas Under State Versus Local Control Related to Balanced Assessment Systems

State Controlled	Locally Controlled
<ul style="list-style-type: none"> <li>• State content and performance standards</li> <li>• Federally required state summative assessments</li> <li>• School accountability systems (e.g., school ratings or rankings)</li> <li>• Teacher standards, licensure, and recertification</li> <li>• Educator preparation program approvals (initial and ongoing)</li> <li>• Additional state-required assessments (e.g., social studies state assessments, dyslexia screeners, universal screeners) or other state-provided assessments (e.g., optional interims)</li> <li>• State supports, guidance, tools, and/or resources offered to local education agencies around curriculum, instruction, and assessment</li> </ul>	<ul style="list-style-type: none"> <li>• Local curriculum, instruction, and assessment decisions, including school- or district-required assessments and classroom assessments</li> <li>• Selection and implementation of professional learning opportunities for teachers</li> <li>• Local assessment policies and practices (e.g., grading policies and requirements regarding curriculum pacing and scripting)</li> </ul>

LEAs (Marion, 2018; Marion et al., 2019; Shepard et al., 2018). This is not to say that decisions at the state level do not affect local actions—the accountability function and design of state summative assessments can incentivize or undermine local efforts to design and implement balanced assessment systems. However, decisions about curriculum, teaching and learning priorities, classroom assessment strategies, and local measures of student progress fall within the purview of local school boards, district and school leaders, and teachers.

Decisions made at the LEA level can have a significant impact on teachers' classroom activity systems (see Chapter 4 of this volume, "Classroom Activity Systems to Support Ambitious Teaching and Assessment"). Over-testing in schools is often due to local testing requirements, although the blame is often laid on the state (Marion et al., 2019).<sup>5</sup> For example, LEAs may require teachers to administer universal screeners, literacy assessments, benchmark assessments, and/or commercial interim assessments beyond those required by the state to track student progress and gauge proficiency. These LEA-selected and -required assessments, in addition to school and teacher assessment preferences, can cause over-testing at the local level, as well as an overreliance on standardized measures of student performance to inform educators' instructional practice. While the prevalence of over-testing is true for all students, it is even more serious for English learners, who often need to take both state- and locally mandated English proficiency tests in addition to all other assessments.

LEAs also control the selection and implementation of professional learning opportunities for teachers. Designing and implementing balanced assessment systems requires educators to have strong pedagogical content knowledge and assessment literacy. Therefore, LEAs need to provide related opportunities for professional learning and capacity-building resources to encourage improved implementation of high-quality local assessment systems (in this volume, see Chapter 5, "Assessment Literacy and Professional Learning," and Chapter 6, "District and School Practices and Assessments to Support a Learning-Centered Vision"). These professional learning opportunities may come from a variety of sources, including regional laboratories and support structures, external professional development providers, and state-provided training.

### State Control

As shown in Table 7-1, states have control over a broad array of factors that can positively or negatively shape local systems of assessment, the quality of the information they yield, and how the information is used. The decisions a state makes regarding many of these factors—particularly additional state-required assessments and state supports, guidance, tools, and/or resources—will depend on its capacity, as well as its vision for teaching, learning, and assessment.

State content and performance standards, federally required state summative assessments, and school accountability systems are constrained by federal statutory regulations and guidance, but the state still has considerable influence over the standards, assessment, and accountability landscape. Specifically, the state—via content

---

<sup>5</sup> One caveat to this statement is the movement of some states toward mandating universal screeners, literacy assessments, or other types of assessments multiple times a year. We express caution about this movement in the section of this chapter titled "State Control."

standards—specifies what is taught, at which grade levels, and at what level of cognitive rigor for all core subjects. Similarly, the annual state test serves to evaluate how successful schools have been in supporting student attainment of identified standards. While federally mandated state summative assessments must meet certain technical requirements, the state has latitude in test design, reporting and administration features (e.g., item types, cognitive rigor, test length), performance standards, and whether to test grades and subjects beyond federal requirements. The state evaluates schools in accordance with the rules of its accountability system. State tests and the associated accountability rules can have an outsized role in shaping local teaching practices and curricula, both positively and negatively.

Because states control teacher standards, licensure, and recertification, SEAs could require educators to meet state-defined assessment literacy requirements to receive a teaching license. Similarly, concerning recertification, the state could offer continuing education credits, micro-credentialing, or badging options related to the demonstration of assessment literacy.

The SEA is also responsible for approving educator preparation programs, both initially and on an ongoing basis—a lever it can use to incentivize or mandate coursework and clinical experiences consistent with the state’s theory of action around balanced assessment systems. We discuss what this type of professional learning might entail and how the SEA could support such efforts later in this chapter.

State boards of education and legislatures decide what, if any, state-defined assessments must be implemented beyond those required by federal law. Examples include K–2 literacy screeners, universal screeners, social studies state assessments, high school end-of-course exams, and additional science testing (i.e., beyond once per grade span). Some SEAs also supply districts and schools with optional interim assessments—either created by the state or purchased through a commercial vendor—that are aligned to the state’s content standards and performance level descriptors. States should be wary of contributing to the possible incoherence and inefficiency of local assessment programs and the over-testing of students when layering additional assessments on top of what is already federally required without a clear theory of action and rationale.

States also determine the types of support, guidance, tools, and /or resources offered to LEAs regarding the implementation of high-quality instructional materials and local assessment practices. Some states provide more support and resources than others due to differences in capacity and vision. In any case, the SEA can play an important role as a convener and connector of LEAs to share best practices. We discuss this issue at length in the section titled “State Action 5: Provide Tools, Resources, and Supports to LEAs.”

## **STATE ACTIONS TO SUPPORT BALANCED ASSESSMENT SYSTEMS**

This section highlights six high-leverage actions SEAs can take in supporting local efforts to design and implement balanced assessment systems (see Figure 7-1). Each action builds from one or more of the state-controlled factors and how the state wants to support locally controlled factors, if at all. Although we acknowledge that it is not yet fully known how these actions interact, we suspect that SEAs must attend to all six actions in some manner to adequately support balanced assessment systems.



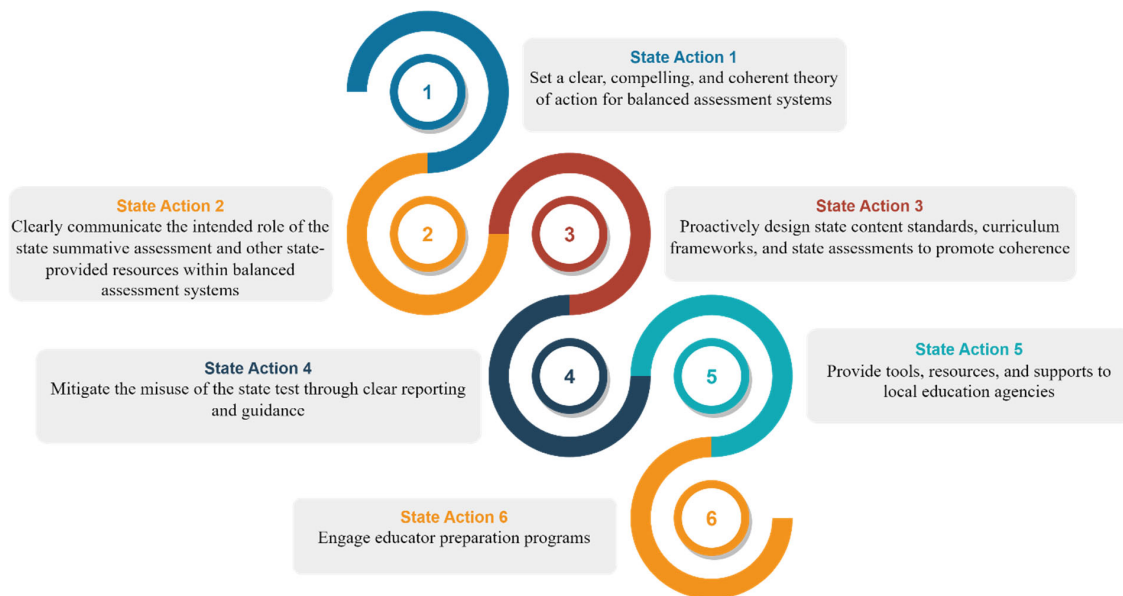


FIGURE 7-1 Six high-leverage state actions to support balanced assessment systems.

### State Action 1: Set a Clear, Compelling, and Coherent Theory of Action for Balanced Assessment Systems

The first action—setting a clear, compelling, and coherent theory of action for balanced assessment systems—is the glue that holds all the high-leverage actions together. A state’s theory of action for balanced assessment systems describes how the SEA understands assessment’s role in supporting teaching and learning, as well as the conditions that must be in place for it to fulfill that role. This theory of action should include assessments required or offered by the state, in addition to those required locally or implemented by educators in the classroom.

Supporting the implementation of balanced assessment systems represents one of several ways a state can champion its educational vision for students. Other avenues include the design of school accountability systems, the development of state policies that influence how and when students learn, and defining course and graduation requirements. To ensure these different approaches represent a thoughtful, coherent strategy rather than a variety of disparate initiatives, a state’s vision should articulate (a) the educational outcomes required to realize the vision (e.g., measures of academic and non-academic performance, participation in extracurricular activities, performance on college and career-ready assessments, acceptance into college or a vocational program upon graduation) and (b) the way those outcomes are likely to be met. Specifically, the theory of action should define the experiences and learning opportunities perceived as necessary for students to achieve these outcomes, as well as the necessary structures, interactions, and information for schools and educators to effectively incentivize and support those opportunities.

A SEA’s theory of action for balanced assessment systems should describe the type of assessment information needed by different stakeholders and how the information

gleaned should be prioritized and utilized to positively impact teaching and learning consistent with the state's vision. Specifically, what information should the state assessment program provide to help achieve the state's educational goals? What information should be generated by other levels of the system (district, school, and classrooms)? How can the SEA help ensure that the state's assessments work together—and not at cross purposes—with district, school, and classroom assessments, supporting rich learning environments?

While SEAs may differ in how they support or enforce key aspects of their theory of action, all SEAs should clearly and consistently describe the role of the state summative assessment. This description should include both how results should be used and how decisions about assessment design are intended to influence what happens in schools. Since the primary purpose of the state summative assessment is to monitor and evaluate school quality, there is no direct link between the information afforded and how to improve teaching and learning practices. The information gleaned from the state summative assessment is too distal from instruction, not at the right grain size, and not timely enough to shape the daily interactions of teachers, students, and the content (Evans & Marion, in press). Furthermore, as discussed in Chapter 4 of this volume, “Classroom Activity Systems to Support Ambitious Teaching and Assessment,” these assessments do not account for the classroom learning environment or consider students' cultural and social backgrounds and funds of knowledge, so they cannot provide information to support ambitious teaching. Future state assessments could be designed to account for learner characteristics, but more personalized and flexible approaches to state standardized assessment do not currently exist (Buzick et al., 2023).

However, the state assessment does provide useful aggregate data for school and district leaders to consider when making programmatic decisions such as how to allocate resources and support, the need for curriculum and staffing modifications, and the effectiveness of new initiatives or programs. In addition, state test data can and should have a positive—albeit indirect—impact on teaching and learning, consistent with the intended role of the assessment in the SEA's theory of action. Table 7-2 reflects two roles and associated theories of action for how a state's summative science assessment design and/or associated resources may indirectly influence teaching and learning.

As shown in these examples, the theory of action reflects an assumption that state assessment design and resources will promote actions or practices that ultimately lead to improved teaching and learning. As with any theory of action, steps must be taken to ensure that these assumptions hold (e.g., that the tools and resources are useful, that educators have time to work together, that sample materials and tasks are high quality) and the desired impact is realized.

Any theory of action that over-emphasizes the role of the state summative assessment program in supporting teaching and learning is bound to cause imbalance. The SEA can support more balanced assessment systems by filling in the missing links from system components such as state content standards and performance expectations to mechanisms that lead to systemic improvement and change. A key aspect of a state's theory of action for supporting balanced assessment systems is therefore clarifying how the state assessment program should inform or work with classroom curriculum, instruction, assessment, and other system components to promote student learning (see also State Action 3 later in this chapter).

**TABLE 7-2** Two Abbreviated Example Theories of Action for State Science Assessment Design and/or Associated Resources

What Is the Role of State Summative Assessment?	How Will the State Assessment Positively Impact Teaching and Learning (Abbreviated Theory of Action)?	What Are the Implications of These Decisions for State Assessment Design and Other Necessary Resources?
<b>Example 1</b>		
To signal the type of authentic, complex tasks students should be able to engage with to demonstrate science learning as envisioned in the Next Generation Science Standards (NGSS).	It will influence how educators engage with students and each other to teach and evaluate the attainment of science expectations within and across years.	Assessment must include one or more high-quality authentic performance tasks, which may impact the time it will take students to complete the assessment and the cost of test design and scoring.
To provide data that help schools and educators evaluate how well existing curriculum and instruction prepared students to generalize their learning to novel tasks.	It will help design curriculum and instruction materials that focus on how to identify and solve authentic problems rather than only the attainment of discrete knowledge and skills.	The state must provide resources (e.g., sample tasks, scoring rubrics) and training that will help educators prepare students for success.
<b>Example 2</b>		
To clarify how the expectations reflected in the NGSS are distributed and evaluated within and across grades given the state’s vision for science education.	It will support schools and districts in establishing a strategy for addressing NGSS expectations within and across grades.	The state must provide resources (e.g., released items) and training that clarifies how the expectations underlying the standards should be addressed and evaluated within and across grades.
To provide data that allows schools and educators to evaluate how well existing curriculum and instruction prepared students to meet expectations at the end of a particular grade span.	It will allow districts and schools to collaborate in the development of curriculum, instruction, and assessment resources based on a shared understanding and trajectory for science attainment.	There is a need to create summative assessment frameworks that complement these resources and reflect priorities for monitoring and evaluating performance (e.g., reporting categories).

If a SEA provides other assessment-related tools, resources, and support—see also State Action 5 later in this chapter—its theory of action should explain how they are intended to support the state’s educational goals. For example, if the SEA provides free assessment literacy resources to all classroom educators and school and district leaders, then the theory of action should explain how providing those free resources is logically connected to improving classroom instruction and assessment practices.

Figure 7-2 provides one hypothetical depiction of a state’s theory of action for balanced assessment systems. The state-controlled assessment components include state-required accountability assessments as well as other state-provided—yet optional—interim and classroom assessment tools, resources, and support (see the yellow box in Figure 7-2). The depiction is meant to communicate how state-required assessments are intended for the limited purpose of program quality monitoring and evaluation. Yet because this hypothetical SEA wants to promote balanced assessment systems, providing interim and classroom assessment tools, resources, and supports can promote and support the quality of local assessments in the teaching and learning feedback cycle. The locally controlled assessment components are in the green box in Figure 7-2.

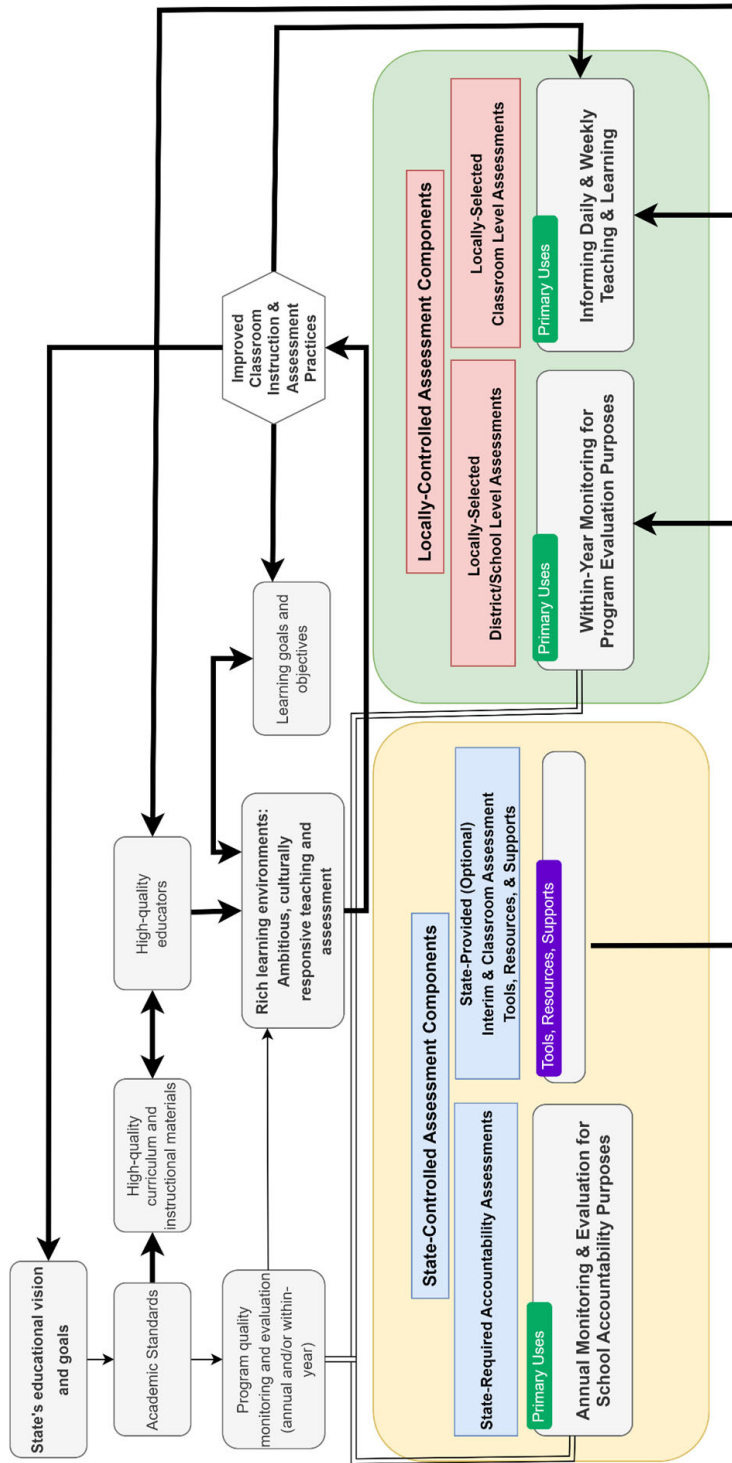


FIGURE 7-2 One hypothetical depiction of a state's theory of action for balanced assessment systems.

It is beyond the scope of this chapter to go into detail about the theory of action, but what should be clear from Figure 7-2 is that there is a complex array of interactions expected and desired among the (a) state-provided tools, resources, and support; (b) locally controlled assessment components; (c) rich learning environments; and (d) improved classroom instruction and assessment practices. The ultimate objective of the theory of action is to support the state’s educational vision and goals for its graduates while recognizing that other offices or departments at the SEA—as well as local educational systems—are also working toward the same vision and goal.

We did not include the hypothetical example to suggest that every SEA needs to provide a similar level of support for locally controlled assessment components. Every SEA has unique concerns, needs, and capacities that drive which goals and problems are prioritized and consequently what solutions are perceived as most likely to support intended outcomes. These unique needs are where identifying general constraints and requirements related to state laws and court rulings, federal laws and regulations, state historical considerations, student and school demographics, fiscal constraints, and/or capacity constraints and limitations is critical, as these constraints directly affect local solutions.

### **State Action 2: Clearly Communicate the Intended Role of the State Summative Assessment and Other State-Provided Resources Within Balanced Assessment Systems**

A SEA can have a clear, compelling, and coherent theory of action for balanced assessment systems and nonetheless fail to communicate it to stakeholders. Communication is the bridge from ideas to implementation—it conveys the rationale for and motivation behind the SEA’s decisions regarding the state summative assessment program and the provision of additional assessment resources (see also State Action 5 later in this chapter). Communication builds awareness and buy-in and educates stakeholders about the meaning and value of balanced assessment systems.

The announced purpose of the state’s summative assessment program is one of the most important messages any SEA can communicate to its stakeholders (e.g., parents, students, classroom educators, school and district leaders, state policymakers, and the public). We have noticed confusion and misconceptions about state summative assessment programs that are due to a lack of accurate messaging. For example, there are abundant mixed messages about how teachers should use state summative assessment results to inform instruction. State leaders, among others, tend to promote the usefulness of their state’s assessments. However, in doing so, they often, if unwittingly, overstate the instructional value teachers can derive from the quantitative results, given the accountability demands and associated design limitations of the assessments (Coburn & Turner, 2011; Datnow & Hubbard, 2015; Evans & Marion, in press; Faxon-Mills et al., 2013).

State leaders must clearly communicate the limitations of the state summative assessment program, particularly for informing classroom instruction. State test information is not useful for informing daily or weekly classroom decision making, including what teaching and learning experiences must be adapted to better meet students’ specific learning needs (Shepard, 2021; Shepard et al., 2018). The timing of



state assessment information, unclear relationship to the enacted curriculum, and grain size of received information (e.g., scale score and achievement level) do not support direct instructional relevance for modifying or adapting teaching and learning practices in real-time (Evans, 2022; Evans & Marion, in press; Faxon-Mills et al., 2013; Marion, 2019b; Shepard, 2021; Shepard et al., 2018).

State, district, and school leaders must clearly and consistently specify the intended purpose and use of state assessments as per federal law—namely, monitoring and evaluating school quality for accountability purposes in a way that is comparable across schools and districts in the state. State test results allow SEAs and LEAs to monitor achievement trends and gaps; examine the efficacy of interventions, programs, and curriculum materials; direct resource allocation; and identify new and promising practices. State test results are useful for these purposes because state tests provide a reliable information source that is comparable over years and is available for every student in the tested grades and subjects.

Connecticut, for example, used its 2022 state test data and National Assessment of Educational Progress scores to identify a state-wide weakness in middle school math performance due to the education interruptions during the COVID-19 pandemic. As a result, the Connecticut State Department of Education designed a competitive grant program for the 2023–2024 school year that creates an intensive mathematics tutoring program for students in Grades 6–9, including funding and a vetted list of approved tutoring providers (The Office of Governor Ned Lamont, 2023). The SEA will then use an education research collaboration it established with institutions of higher education across Connecticut (State of Connecticut, 2023) to monitor the effectiveness of this tutoring program using state test data from participating schools.

Connecticut’s clear communication about the intended purpose and use of the state assessments should also transfer to any state-provided assessment or assessment-related resource curated by the state without additional cost to LEAs (see also State Action 5 later in the chapter). The key point here is that the SEA must articulate to stakeholders why they have provided these tools and resources, their intended uses and users, and how they can support or work against more balanced assessment systems.

As with state assessment information, SEAs bear the responsibility of explaining the intended use of any provided or required assessments. For example, if the intended use is to support program evaluation conducted by school and district leaders, then the state must show the chain of reasoning, assumptions, mechanisms, and professional learning that connects the information gleaned from the assessment to that use. Doing so increases the likelihood that the assessment information will be used to make decisions that provide for better student learning, as well as preventing misuse, incoherence, and over-testing.

### **State Action 3: Proactively Design State Content Standards, Curriculum Frameworks, and State Assessments to Promote Coherence**

A SEA does not promote balanced assessment systems as an end in and of itself; rather, many SEAs want to support and incentivize a robust vision of teaching and learning, academic achievement, and inclusive educational practices for all students.

A SEA can use the levers within its control to further this vision, including supporting the design and implementation of rigorous content standards; curriculum frameworks; state test designs with appropriate accessibility features; and other tools, resources, and support that support high-quality local curriculum, instruction, and assessment practices.

For example, one input from the SEA into balanced assessment systems is the state summative assessment program. States have considerable latitude in terms of how they design their state assessment program, as long as the tests meet federal peer review requirements. One aspect of meeting federal peer review requirements is to demonstrate that the assessment adequately represents the depth and breadth of the state content standards. Consequently, the design of the state content standards and associated curriculum frameworks can help promote coherence among curriculum, instruction, and assessment at the classroom and state levels. *Knowing What Students Know* (National Research Council, 2001) indicates that coherence is demonstrated when assessments within a system are linked through a clear conceptual base and specification of learning targets. Below, we argue that a SEA can promote coherence by proactively designing state content standards, associated curriculum frameworks, and state tests in a way that signals valued instructional priorities.

#### *Design of State Content Standards and Associated Curriculum Frameworks*

State content standards can shape teaching, learning, and assessment in classrooms because they frame what is important to know and be able to do in a specific content area at the end of each grade. State content standards underlie decisions about the design of the state assessment program and are the basis against which decisions about the quality and appropriateness of local assessments are made (e.g., alignment to standards). State content standards serve as a through line that extends from the state to the classroom and consequently play a large role in ensuring coherence (see Chapter 1 of this volume, “Reimagining Balanced Assessment Systems: An Introduction”). Currently, most state content standards are long lists of discrete knowledge, skills, and understandings that are isolated from other content within and across domains, silent on intended generalization and use, and not developmental. One way that states can promote more coherence and potentially advance impactful use of state assessment data is to write richer learning expectations for students—connecting competencies with other content within and across domains in desired ways, explicitly stating intended generalization and use, and displaying developmental structure and sequences.

States must use research on the way students learn and demonstrate more sophisticated knowledge and expertise within a domain to design state content standards and supplementary documentation and guidance, like curriculum frameworks to support ambitious teaching and equitable assessment practices. For example, the Next Generation Science Standards intentionally tried to reshape curriculum, instruction, and assessment in K–12 science classrooms to better reflect research on the ways students learn science (National Research Council, 2012). Similarly, the Common Core State Standards in math signal to the field that mathematics education is more than just procedural skill and fluency and “build on the best of existing standards and reflect

the skills and knowledge students will need to succeed in college, career, and life” (Common Core State Standards Initiative, 2021).

Some states, including California, have helped teachers understand how to implement content standards by creating curriculum frameworks (California Department of Education, n.d.). These frameworks help teachers faced with lists of discrete standards understand how to cluster and group the standards for instructional purposes, such that teachers can focus on the big ideas that are important at each grade level and recognize the underlying learning progressions tying the standards together. Understanding the underlying progressions supports accelerated learning and other differentiated approaches to instruction because it gives teachers both a heuristic for interpreting evidence of student learning and knowledge about what instructional moves are most likely to help students progress toward proficiency.

### *State Test Design*

State tests serve as examples of inclusion practices both from a participation perspective (every student in federally required grades and subjects is counted in participation rates) and from an accessibility perspective (every student can show what they know and can do). Additionally, SEAs operationalize and demonstrate their values about how students learn a discipline and what instruction should be prioritized through the design or adoption of state tests (National Research Council, 2001, 2003). Ideally, states should design or adopt—in the case of assessment consortia—their state assessment program with a clear understanding of (a) how state tests will promote the instructional priorities that the state values and wants to see implemented; and (b) an understanding of how the state summative assessment should complement information collected through local assessment systems. Furthermore, states should strategically engage a diverse array of stakeholders in the assessment design and specification process, including those who represent the cultural, ethnic, racial, and special populations present in the state. Involving stakeholders from the beginning of the state assessment design process helps ensure the cultural validity of assessment results (Shultz & Englert, 2021) and models the type of stakeholder engagement and inclusivity desired at the local level.

States that are part of assessment consortia such as Smarter Balanced will have additional layers of complexity to consider. For example, assessment consortia, by design, somewhat constrain individual state decisions because a set of distinct state testing programs is replaced by one collective consortium testing program. States may have some leeway to adjust the test blueprint and reporting structure, but there are limits to what they can personalize when they are part of consortia.

Per State Action 1, presented earlier in this chapter, the state assessment program should be designed with a clear understanding of how it should support better educational decision making—ultimately supporting student learning—and the influence it is likely to have on local curriculum, instruction, and assessment practices. For example, part of the design work for a state assessment is considering tradeoffs associated with different test design features including item types, content priorities, adaptive testing models, and length of the test. Including longer constructed-response items and performance-based tasks on state assessments can signal the importance of cognitively

rigorous teaching, learning, and assessment. However, adding more complex item types usually results in a longer test and more expensive scoring. To promote coherence, the design of the state tests should send a clear and consistent message about what is important for teachers to teach and students to learn, and at what level of cognitive rigor.

Figure 7-3 shows a released item from the Spring 2022 Grade 10 Mathematics state test in Massachusetts (Massachusetts Department of Elementary and Secondary Education, 2023). Figure 7-4 shows an Algebra I performance task from the Mathematics Assessment Project (Mathematics Assessment Resource Service, 2011). Imagine a state test that contains only selected-response items like those found in Figure 7-3 versus a mixture of item types, including performance tasks like the one found in Figure 7-4. What view of human learning and development (see Chapter 3 of this volume, “Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems”) and associated classroom activity systems (see Chapter 4 of this volume, “Classroom Activity Systems to Support Ambitious Teaching and Assessment”) do these different test designs support or constrain? Our point here is simple: SEAs can harness the signaling function of the state test to either promote or work against its vision of teaching and learning.

Whatever decisions are made by the SEA about the design of the state test should be clearly communicated to LEAs through blueprints, guidance documents, item samples, released annotated items, and practice tests, among other resources. To ensure that the state assessment program and associated resources reflect the state’s theory of action, SEAs must clearly articulate their requirements when they release a request for proposals to vendors, including the specific claims, interpretations, and uses the state summative assessments and any additional state-provided assessments must be designed to support. Test vendors are responsive, not vision-casting entities. A vendor’s job is to design assessments that reflect the state’s goals, vision, and theory of action—so these must first be defined by the SEA. This is especially relevant if a SEA is looking for its state assessment program to include novel elements.

In addition, state tests should be designed, and achievement levels set, based on realistic and attainable performance expectations. Realistic and attainable expectations are essential if state tests are to have a positive influence on educational decision making and student learning. Unattainable expectations can undermine motivation and encourage inappropriate test preparation and use.

What is the solution of this equation?	$3(x + 5) = 5x - 7$
A. $x = -1$	
B. $x = 4$	
C. $x = 6$	
D. $x = 11$	

**FIGURE 7-3** Selected-response item from MCAS 2022 grade 10 mathematics released items. SOURCE: Massachusetts Department of Elementary and Secondary Education (2023).

On the grid are eight points from two different functions.

A certain linear function passes through exactly four of the points shown.

A certain quadratic function passes through the remaining four points.

For the **linear** function:

1. Write the coordinate pairs of its four points.

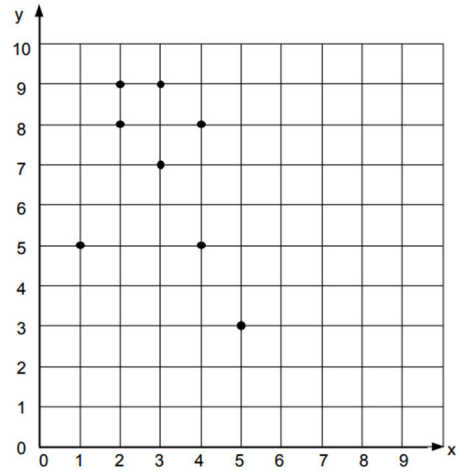
---

---

---

---

Draw the line on the grid.



2. Write an equation for the function.

Show your work.

---

For the **quadratic** function:

3. Write the coordinate pairs of its four points.

---

---

---

---

Draw the graph of the function on the grid.

4. Write an equation that fits the quadratic function.

Show your work.

---

**FIGURE 7-4** Algebra I performance task from the Mathematics Assessment Resource Service Task Bank.  
SOURCE: Mathematics Assessment Resource Service (2011).



#### **State Action 4: Mitigate Misuse of the State Test Through Clear Reporting and Guidance**

State test misuse can unbalance assessment systems at the local level. For example, state test results are sometimes inappropriately used to screen students out of algebra or place students into remedial coursework or non-flexible groups or tracks. Similarly, while state test results provide highly reliable information that could be used to monitor school improvement efforts over time, some districts and schools administer additional assessments for the same purpose because local users are not aware of how to use state assessment results in this way. States can mitigate state test misuse—or lack of use—through clear reporting guidance directed to users of the assessment system. These mitigation strategies promote the utility and efficiency of the entire system, which are key characteristics that support balance.

##### *State Test Reporting Features and Guidance for Interpretation and Use*

Reporting is often an afterthought or post-hoc activity following assessment design and data collection. And yet, how state assessment information is communicated to stakeholders—parents, students, classroom teachers, school and district leaders, and the public—and ultimately interpreted and used can significantly impact how stakeholders engage with, perceive, or value information from the state test or other state system components. Balanced assessment systems should provide their many diverse stakeholders with the information they need to make timely, accurate educational decisions that ultimately support student learning. Score reports are the vehicle for communicating the test’s results to stakeholders and therefore must be crafted with care.

To support the utility of state information, score reports and associated resources should be designed with specific users in mind. Currently, many state-produced reports lack clear user guidance (e.g., suggested actions for school and district leaders, classroom educators, or parents), which could leave system users to interpret and use assessment results inappropriately. However, the adequate and appropriate interpretation of test scores—let alone moving from interpretation to actionable next steps—requires a high level of assessment literacy, time, and effort. Ensuring adequate and appropriate interpretation of scores might be better met, and time and effort better spent, if states provided a selection of high-impact reports that presented student and aggregate test results in multiple ways with a few high-leverage actions that different system users could take based on the results. For school and district leaders, these high-leverage actions might include gathering more contextual information on program implementation and teacher curriculum supplementation. The state likely has historical data on students, schools, and districts. This information could be used to create a reporting system that provides district- and school-level reports summarizing trends in overall student performance and reporting category for each grade and content area. This type of reporting system could help save local leaders’ time, as they might otherwise try to create these reports on their own. Additionally, the reporting system might propose questions for the LEA to investigate based on student performance and trends over time.

Additionally, as with test design, states demonstrate their values through choices reflected in the state reporting system. Score reports and other assessment-related

guidance often use reporting structures and deficit-based labels—for schools and students—that can lead to interpretations that promote systemic inequities. For example, labels used to categorize student performance as ‘failing’ or ‘well below proficient’ can influence perceptions of ability and consequently teaching practices in ways that contradict rich and inclusive learning environments.

### **State Action 5: Provide Tools, Resources, and Support to LEAs**

As stated throughout this chapter, states have little, if any, control over the composition and implementation of local assessment systems. However, SEAs can influence and promote high-quality assessment policies and practices at the local level by directly creating or curating tools, resources, and support for assessment-related endeavors. SEA personnel can further their work in these areas by attending professional conferences and meetings, as well as engaging in professional networks where they can learn from other SEAs, researchers, practitioners, and organizations. We discuss tools, resources, and support created by SEAs as one of the final state actions because they must flow from the state’s vision, associated communication strategies, and other proactive and mitigation activities related to the state assessment program.

In this section, we discuss the assistance states should provide LEAs to some degree to inform the design and implementation of balanced assessment systems at the local level. However, we recognize that SEA and LEA capacity will influence how SEAs attend to these suggestions. We do not expect that all SEAs will create or compile the same set of tools, resources, and/or support, or provide the same set of supporting elements to all LEAs, but the five categories in Figure 7-5 should be considered by all SEAs as high-leverage opportunities to assist LEAs.

As shown in Figure 7-5, we organize the types of tools, resources, and support SEAs can provide into five categories: (1) curriculum and instructional material reviews; (2) a professional learning provider clearinghouse; (3) local assessment practices support; (4) local assessment system auditing tools; and (5) assessment literacy resources. The actions within each category are listed in order of those that require the least to greatest amount of state capacity, involvement, and effort. Although these categories are listed separately in Figure 7-5 and the sections that follow, they are also interrelated.

#### *Curriculum and Instructional Material Reviews*

High-quality curriculum and instruction are central to ensuring that all students have access to grade-level, standards-aligned teaching and learning experiences. It is a fundamental equity issue that all students have the opportunity to learn what students statewide are supposed to know and be able to do by the end of each school year. However, the implementation of high-quality curriculum and instructional materials is not occurring in many U.S. classrooms (Kaufman et al., 2020). Instead, many teachers spend inordinate amounts of time supplementing their curriculum for different reasons and with largely unknown effects (Silver, 2022).

High-quality curriculum, instruction, and formative assessment processes—aligned to the content and cognitive complexity of the state’s content standards—are the mechanisms by which student learning improves. These factors are especially relevant

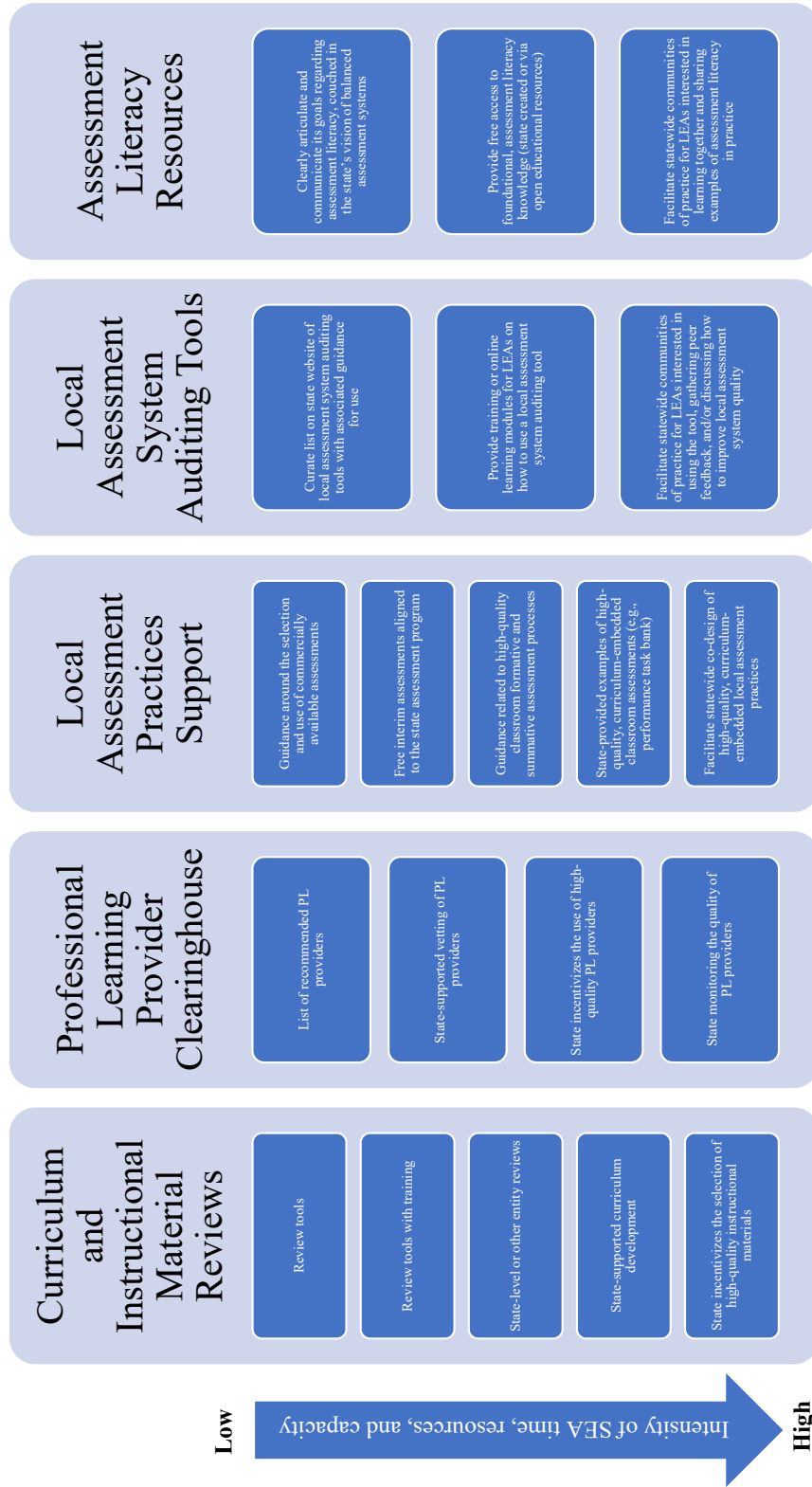


FIGURE 7-5 State tools, resources, and support for LEAs related to balanced assessment systems.

to balanced assessment systems because local assessments should not move students away from high-quality instructional materials—assessments should cohere with and mutually support high-quality teaching and learning experiences within the curriculum. *Ensuring that all LEAs understand what constitutes high-quality instructional materials (HQIMs), including using criteria to evaluate the quality of the curriculum-embedded assessments and assessment processes within those materials, is arguably the most important educational action a state can take to facilitate the design and implementation of balanced assessment systems at the local level.*

Polikoff (2021) convincingly argues that the failure of standards-based reform is due, in large part, to decentralized governance structures that result in poor and inequitable standards implementation. Local control over curriculum results in very little standardization of common curriculum materials across and within states, and there are real differences in curriculum quality that create systemic inequities in students' opportunities to learn. Polikoff argues that radical change is needed to improve instruction at scale and that SEAs have a key role to play in this change process—especially in providing more oversight and support related to HQIMs. One initial action step, Polikoff notes, is for states to collect good data on what curricula are being implemented in classrooms.

Although states have limited control over local curricula, they can provide a variety of related support to facilitate high-quality standards implementation. For example, in 2017, the Council of Chief State School Officers (CCSSO) and a cohort of 12 interested states launched the High-Quality Instructional Materials and Professional Development (IMPD) Network, “dedicated to ensuring that every student, every day, is engaged in meaningful, affirming, grade-level instruction” (Council of Chief State School Officers, n.d.). The IMPD Network provides guidance and case studies to SEAs showing how they can engage with LEAs to adopt HQIMs and ensure access to professional development opportunities that are aligned with those materials (Council of Chief State School Officers, 2021a, 2021b, 2022b, 2022c, 2022d). A recent RAND study on the states in the IMPD Network showed that the incentives the states are using have been effective, especially in mathematics, to create higher rates of adoption and use of standards-aligned HQIMs (Doan et al., 2022).

However, not all states have the same flexibility and capacity. For example, the Wyoming Department of Education operates under a legislative mandate that requires the SEA to remove itself from all local curriculum decisions due to concerns of state overreach. Due to this state statute, the Wyoming Department of Education does not have an office of curriculum and instruction. Therefore, because the provision of state support for HQIMs will be influenced by the size and capacity of the state department of education, as well as a state's legislative freedom around local curriculum, our recommendations below fall along a continuum of state-level involvement.

The five SEA actions that could influence curriculum and instructional material reviews are listed below in order from the least to most required state involvement:

- **Provide high-quality curriculum and instructional material review tools.** The state could focus on signaling the quality of instructional materials to LEAs using HQIMs review tools (Council of Chief State School Officers, 2022b, 2022d). States could also design their own review tools. Additionally, they could adopt or adapt

- existing tools, such as EdReports' Curriculum Review Tools (EdReports, 2022b), or state-developed curriculum review rubrics such as those found in Louisiana, Massachusetts, Mississippi, New Mexico, Rhode Island, Tennessee, and Texas (Council of Chief State School Officers, 2022d). It is especially important for states to ensure that any review of curriculum materials interrogates the quality of the curriculum-embedded formative assessment processes and summative assessments to support more balanced assessment systems.
- **Provide training on how to use and apply the review tools.** In addition to providing review tools, SEAs could provide training to help LEA leaders fully understand the tools and practice conducting sample reviews. The training could be self-paced and accessible on demand or provided through in-person workshops. If the state adopts existing review tools, it could promote the corresponding training or certification (EdReports, 2022a). These trainings could also show LEAs potential solutions if they identify gaps in the curriculum—for example, how supplementary materials can be added to achieve more robust curriculum and standards implementation.
  - **Conduct state-level review of curriculum and instructional materials (or link to other entity reviews).** Some states directly review curriculum materials. For example, Louisiana has an instructional materials review process where the state provides annotated reviews of K–12 curriculum materials in ELA, math, science, and social studies using evaluation criteria in the state's review tools (Louisiana Department of Education, n.d.). A review produces one of three rankings (Tiers 1–3), reflecting the degree of alignment with the state's content standards and vision of teaching and learning. Although each Louisiana school system can decide whether to draw on these reviews, it is in their best interest to do so insofar as state funding is tied to the selection of Tier 1 curriculum materials. Other states could follow Louisiana's template or decide to follow their own state-level review process, which would make choosing the appropriate curriculum much easier for LEAs. However, the ongoing review of curriculum materials at the state level entails considerable work. Consequently, some states could take advantage of reviews provided by others, such as EdReports (2022b) for English language arts (ELA), math, and science. The state could also consider implementing an EdReports review that is specific to their state (e.g., Arkansas EdReports).
  - **Create state curriculum and instructional materials with no adoption requirement.** States could also decide to create their own curricula and offer it to LEAs with no adoption requirement. For example, Louisiana educators have produced K–12 ELA Guidebooks for Louisiana students, which most of the state's school systems use for their ELA curriculum and is offered free of charge to Louisiana school systems. Alternatively, states could partner with an open educational resource curriculum provider to create free, high-quality curriculum materials for their school systems' consideration. For example, 10 states currently partner with OpenSciEd for just this purpose.<sup>6</sup> One advantage of state involvement in curriculum development is that the state can fold its vision for balanced assessment systems directly into curriculum design. For example,

---

<sup>6</sup> See <https://www.openscienced.org/why-openscienced/partner-states>.



- the state could create a K–12 curriculum and instructional materials that are accompanied by high-quality formative assessment processes and curriculum-embedded classroom assessments. The practice of developing a tailored curriculum holds considerable promise for disseminating the foundational knowledge necessary to support best practices in classroom assessment.
- **Incentivize the selection of high-quality instructional materials.** The four previous proposed SEA actions focused on states signaling the quality of instructional materials. The final, most time-intensive, but also most impactful action, would be for a state to incentivize the selection of HQIMs. Some states accomplish this by establishing financial incentives for districts that select materials from the state’s recommended list of HQIMs. These financial incentives could include state competitive grants, school improvement funding, requirements for use of some federal funds, COVID-19-pandemic-related federal relief funding appropriations, and statewide contracts for HQIMs that reduce the cost of the materials (Council of Chief State School Officers, 2022d).

#### *Professional Learning Provider Clearinghouse*

In the same way a state can review curriculum materials, it can also increase the number of teachers who have access to high-quality professional learning about curriculum and standards implementation by incentivizing a strong vendor marketplace. CCSSO’s IMPD Network created a guidance document that describes four different ways a SEA can support districts in using high-quality professional learning providers (Council of Chief State School Officers, 2022a):

- provide districts with a list of recommended professional learning providers;
- support districts in vetting professional learning providers;
- incentivize the use of high-quality professional learning providers; and / or
- monitor the quality of professional learning providers.

For example, Louisiana reviews professional learning vendors who target core academic subjects and then provides a vendor guide to all Louisiana school systems.<sup>7</sup> Louisiana then incentivizes the use of high-quality professional learning providers in the Louisiana Super Application, which is an integrated application Louisiana LEAs use to apply for Title I, Title II, and School Improvement funds every year. Additionally, Louisiana has developed a tool to track professional learning provider use and monitors professional learning quality using LEA and Teacher Satisfaction Surveys. The review and dissemination of high-quality professional learning that supports standards implementation, particularly the assessment-related aspects of that training, is an important step in supporting the implementation of balanced assessment systems. The authors recommend reading *Guidance for States on Supporting District Use of High-Quality Professional Learning Providers*, from the Council of Chief State School Officers (2022a) for examples from other states including Delaware and Rhode Island.

---

<sup>7</sup> See <https://www.louisianabelieves.com/academics/curriculum>.

As curriculum, instruction, assessment, and state content standards should work coherently together to support balanced assessment systems, these professional learning opportunities should also provide training and coaching to skillfully utilize curriculum-embedded assessments within HQIM and to interpret resulting assessment information. These professional development offerings should include evidence-based implementation features such as ongoing job embedded training and coaching, active learning, teacher collaboration, and involvement of school leaders (Shapovalov & Evans, 2022).

### *Local Assessment Practices Support*

Although states do not control local assessment practices and policies (see Table 7-1), they can indirectly support these activities by providing classroom and interim assessment system tools, resources, and support. High-quality local assessment practices must be in place to support the implementation of balanced assessment systems, as the assessment system is composed mainly of these local assessments. Despite its outsized impact, the state assessment program is a small aspect of a balanced assessment system and has a very particular purpose and intended use. These state-provided tools, resources, and support would be optional, and could include:

- providing guidance around the selection and use of interim assessments;
- providing free interim assessments aligned to the state’s theory of action for balanced assessment systems;
- providing guidance related to high-quality classroom formative and summative assessment processes;
- providing examples of high-quality, curriculum-embedded classroom assessments (e.g., performance task bank); and/or
- facilitating statewide support for the co-design of high-quality, curriculum-embedded local assessments.

The first two bullets focus on interim assessments. Interim assessments are optional parts of balanced assessment systems (Marion, 2019a; Marion et al., 2019), although they are ubiquitous and unlikely to fade from use in the near future. Interim assessments are defined as:

Assessments administered during instruction to evaluate students’ knowledge and skills relative to a specific set of academic goals in order to inform policymaker or educator decisions at the classroom, school, or district level. **The specific interim assessment designs are driven by the purposes and intended uses**, but the results of any interim assessment must be reported in a manner allowing aggregation across students, occasions, or concepts. (Perie et al., 2009, p. 6; emphasis added)

The importance of specifying the purposes and intended uses of interim assessments is emphasized in the definition above because interim assessment designs differ and they do not provide the information local educators might need or want equally well (Gong, 2019). There is a strong desire among many educational leaders to procure interim assessments and administer them two to three times over the school year to

gather within-year information on student academic achievement that appears more objective than locally created measures. The state should make decisions on how to support LEAs regarding interim assessments based on its theory of action for balanced assessment systems and the needs of schools and districts. This support could include providing advice about the pros and cons of various interim assessment designs—whether they are commercially purchased or state-provided. It could also include encouraging LEAs to use a thoughtful procurement process that includes specifying use, identifying desired assessment features, and evaluating the technical quality of the interim assessment options (Landl & Lyons, 2023). For example, many LEAs look to commercial interim assessments with the desire to “inform instruction” throughout the year but end up purchasing assessments that are designed to closely mimic the state test design (e.g., NWEA MAP, Renaissance STAR). Unfortunately, local leaders may fail to realize that these types of tests “typically lack sufficient ties to curriculum and instruction to make it possible to provide feedback that leads to improvement” (Shepard, 2005, pp. 2–3). SEAs can provide guidance around interim assessments that helps cut through confusion and marketing claims. SEAs can also decide if providing free interim assessments designed to support specific purposes is important for supporting their theory of action around balanced assessment systems (or not).

The last three bullets above relate to support for high-quality classroom assessment practices. Given rich learning environments foster changes in interactions among the teacher, students, and content, states may want to support the conditions for improved classroom instruction and assessment practices. Classroom assessment tools, resources, and support can range from guidance around best practices to state-provided examples of curriculum-embedded classroom assessments to facilitating statewide gatherings of educators to co-design classroom assessments. These example actions are not mutually exclusive, as a state could support all or only one. For example, the Hawai‘i Department of Education is using two recent Competitive Grants for State Assessment awards to design and implement state-provided, optional classroom assessment tools, resources, and support (U.S. Department of Education, 2022). These optional resources focus on increasing the quality of classroom assessments and assessment processes by creating a bank of exemplar performance tasks, tied to the curriculum, with related instructional guides (Hawai‘i Performance Assessment Task Bank, 2022). These activities and outputs are intended to build educator capacity to create rich learning environments and ultimately advance student learning through improved classroom instruction and assessment practices.

#### *Local Assessment System Auditing Tools*

Part of a state’s communication strategy should include sharing its vision of balanced assessment systems with LEAs (see also State Action 2)—but vision sharing is not enough. Rather, LEAs need tools to help them understand what balance means in practice, as well as tools for auditing and evaluating the balance of their local assessment systems (i.e., state-required annual achievement testing, school- and/or district-required assessments, and classroom assessments).

Local assessment system auditing tools and resources can help educators at all levels reflect on the relevance, usefulness, coordination, and quality of the set of assess-

ments that their local assessment system comprises, including state, district and school, and classroom assessments (Chappuis et al., 2016). Assessment audits can help educational leaders identify who needs assessment information when and for what purpose to evaluate the quality of their local assessment system—eliminating inefficiencies, redundancies, and low-quality assessments in the process.

States can support the use of local assessment system auditing tools and resources in several ways:

- curate list on state website of local assessment system auditing tools with associated guidance for use;
- provide training or online learning modules for LEAs on how to use a local assessment system auditing tool; and/or
- facilitate statewide communities of practice for LEAs interested in using the auditing tool, gathering peer feedback, and/or discussing how to improve local assessment system quality.

The quality of local assessment systems is critical because previous analysis has shown that a majority of assessment burden and over-testing arises from locally required assessments (Lazarin, 2014). Local assessment system audits can help promote more balanced assessment systems by evaluating and analyzing the assessments administered in the district or school based on intended users and uses of the assessment information. In other words, is the system of assessments providing the necessary information for specific users to make educational decisions that support student learning at the right time and the right level of specificity and relationship to the enacted curriculum?

Auditing resources (Coladarci, 2002; Council of Chief State School Officers, 2015) and auditing tools for local assessment systems are available for use (Achieve, 2014; Chappuis et al., 2016; EducationFirst, n.d.; Evans & Thompson, 2022b, 2022c; Martineau et al., 2018). The Georgia Department of Education, for example, partnered with the Georgia Partnership for Excellence in Education to pilot how to help school districts test “smarter” rather than more often (Georgia Partnership for Excellence in Education, 2022). Most of these auditing tools require broad stakeholder engagement and provide a comprehensive framework for thinking about local assessment system quality. All of these tools involve time and effort—some more so than others. Only some of these tools include audits of classroom-level assessment systems, as well as state-, district- and school-level. In most cases, the quality of each assessment is not a focus of systems-level evaluations but could be a follow-up or concurrent activity.

Auditing tools need not be overly complex or involve all potential stakeholders to be effective. However, auditing tools should reflect the complexity of systems and the range of students who participate in them (e.g., types and frequency of information needed to support students with disabilities or English learners). The interaction among state-, district-, school-, and classroom-level assessments is important because state assessments may provide information that overlaps with the assessment needs at the district or school level, providing duplicative and redundant information. Without analyzing the assessments altogether, those tasked with auditing and evaluating the system would not see the overlaps and redundancies. The complex web of assessments

must be analyzed together to evaluate the extent to which the entire system exhibits the features of balanced assessment systems.

### *Assessment Literacy Resources*

Assessment literacy encompasses the knowledge and skills that educators (both classroom teachers and school and district leaders) need to appropriately utilize assessments to inform educational decisions about student learning (Stiggins, 1991) (see also Chapter 5 of this volume, “Assessment Literacy and Professional Learning”). The SEA’s role in supporting educators’ assessment literacy falls on a continuum of involvement. At a minimum, a SEA should clearly articulate and communicate its goals to LEAs regarding educator assessment literacy and how those goals are couched in the state’s vision of balanced assessment systems.

After this baseline responsibility, SEAs will have differing amounts of personnel and capacity to support assessment literacy initiatives. In light of the various capacities of SEAs, we are not suggesting that every SEA should create its own assessment literacy professional learning resources. Rather, the SEA can link to free resources, such as webinars and self-paced modules, that have been created by other entities. For example, the Michigan Assessment Consortium provides free resources and tools on its website, and the Center for Assessment provides a set of open-access teacher and leader professional learning modules (Evans & Thompson, 2022a).

The purpose behind a SEA providing free access to foundational assessment literacy knowledge is to create and support school conditions that will promote student learning. Assessment, when it is working as intended, provides feedback loops to students and educators that can be used to adjust teaching to the benefit of student learning. The goal should be to establish a common and sufficient level of assessment literacy knowledge and skill for district-, school-, and classroom-level educators so that they can foster best practices in assessment, student learning, and professional collaboration (DeLuca et al., 2019; Xu & Brown, 2016).

The *Standards for Teacher Competence in Educational Assessment of Students* describe what all teachers should know and be able to do related to assessment in their classrooms (American Federation of Teachers et al., 1990). Others have built on these foundational educator assessment literacy standards (e.g., Klinger et al., 2015; Michigan Assessment Consortium, 2016). There are also assessment textbooks for teachers often used in educator preparation programs (Brookhart & Nitko, 2019; McMillan, 2021). Xu and Brown (2016) synthesize that body of work just listed (and more) and delineate assessment literacy foundational knowledge as disciplinary knowledge and pedagogical content knowledge; knowledge of assessment purposes, content, methods, grading, feedback, and peer- and self-assessment; assessment interpretation and communication; and assessment ethics.

SEAs will likely want to identify what various users need to know and be able to do related to assessment as they consider how to support better educational decision making focused on student learning. For example, district and school leaders should possess literacy about assessment commensurate with their respective roles in the educational system. School leaders are more involved in teacher supervision and instructional coaching, so assessment literacy related to classroom assessment processes



is crucial. District leaders are more involved in making decisions about resource allocation based on test scores, purchasing interim assessments, and setting grading policies so assessment literacy related to those topics is important. Teachers, on the other hand, need training and coaching around high-quality formative and summative classroom assessment processes (see Chapter 5 of this volume, “Assessment Literacy and Professional Learning”).

Another way a SEA can support assessment literacy is to facilitate statewide convenings or communities of practice for LEAs that are interested in learning together and sharing examples of assessment literacy in practice. Arranging these convenings will require more involvement from the SEA if the SEA will also be serving as a facilitator, but over time the state could train others to take a leadership role or run more informal networks of support.

To sum up State Action 5, as SEAs increasingly integrate their actions related to curriculum and instructional material reviews, a professional learning provider clearinghouse, local assessment practices support, local assessment system auditing tools, and assessment literacy resources, it is also increasingly likely that the SEA’s efforts will lead to systemic and scalable change. SEAs and LEAs must work together as partners to accomplish school reform. It may be that the most important role of the SEA is to serve as a convener and collaborator for local leaders and classroom educators to work together toward lasting education reform.

### **State Action 6: Engage Educator Preparation Programs**

States control teacher, principal, and superintendent standards, licensure, and recertification. SEAs also approve educator preparation programs to ensure that the teachers who graduate from these programs are highly qualified and well prepared to serve all students effectively. We are not suggesting that the solution to ensure all incoming teachers are assessment literate is for SEAs to mandate more coursework for teacher and school and district leader preparation. Rather, we are suggesting that SEAs provide guidance to educator preparation programs about how to integrate and embed assessment literacy principles within core coursework.

States could use their role in the educator preparation program approval process to ensure that the programs provide the coursework and clinical training necessary to support the state’s vision and theory of action related to balanced assessment systems. In particular, states could ensure that teachers and school and district leaders who graduate from the state’s educator preparation programs understand the importance of HQIMs and have the assessment literacy and content knowledge necessary to support coherence among curriculum, assessment, instruction, and the state’s content standards. For example, CCSSO’s IMPD Network provides guidance for state policies related to educator preparation to support HQIMs implementation (Council of Chief State School Officers, 2020). This guidance recommends that SEAs use their policy levers—statewide teacher competencies, initial and ongoing program approval or accreditation, and licensure and recertification requirements—to encourage educator preparation programs to revise their coursework and clinical training experiences to align with the state’s vision and theory of action. Another example of a state-educator preparation program partnership is the HQIMs labs established between the Arkansas

Department of Education and educator preparation programs in the state (Arkansas Division of Elementary and Secondary Education, 2023), which support the design and implementation of balanced assessment systems so long as the preservice training coherently embeds assessment literacy training within the instructional content and methods training.

## CONCLUSION

While many of the decisions that impact local assessment system design and associated policies and practices are made at the district, school, and classroom levels, states play a significant role in promoting the design and implementation of balanced assessment systems. The most important role the state can play in promoting balanced assessment systems is to create and support the right structures and conditions for district and school leaders and classroom educators to be able to do their jobs effectively, thereby improving student learning. States control specific aspects of the education system that can be leveraged into actions to support the right structures and conditions to promote balanced assessment systems. These actions stem from a clear, compelling, and coherent theory of action to achieve a balanced assessment system and include strategic communications, proactively addressing and mitigating issues concerning state assessment programs, and providing LEAs with tools, resources, and support needed for design and implementation of balanced assessment systems at the local level.

Specifically, SEAs should model behaviors, create conditions, and incentivize or facilitate actions that support local efforts to identify or develop assessment tools and practices that provide a comprehensive, coherent, and useful profile of information about student achievement and growth to educators and parents. In this chapter we argue that a SEA can have a significant positive impact on assessment practices at all levels of the educational system by focusing on what it can control and where it has the greatest influence:

- the design of the state’s summative assessments, content standards, and curriculum frameworks;
- the implementation of policies that influence or mandate the use of state summative assessment results beyond those that are federally required;
- ensuring clear communication about the intended purpose and use of state summative assessments; and
- the development of tools and resources that provide support consistent with stakeholders’ needs and intended role in advancing balanced assessment systems.

At the center of these efforts is a clear vision for teaching and learning and a theory of action that clarifies how assessments prioritized at different levels of the educational system should work together to support this vision. The theory of action should clarify the type and range of information that different stakeholders need to support decision making, the role of the SEA and other stakeholders in ensuring the collection and appropriate use of assessment information, and the necessary conditions for stakeholders to fulfill their intended roles. Articulating this vision is critical for ensuring that the state’s actions are consistent with its theory of action, but even more importantly,

articulating this vision ensures that districts and schools have a clear understanding of what it means to design and implement a balanced assessment system. How the SEA situates the state summative assessment program relative to this goal can positively impact what happens at the district, school, and classroom level if the vision is clearly communicated and reinforced through the development of tools, guidance, and other resources that support local efforts.

## REFERENCES

- Achieve. (2014). *Student assessment inventory for school districts*. <https://www.achieve.org/assessment-inventory>.
- American Educational Research Association. (2015). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, 44(8), 448–452. <https://doi.org/10.3102/0013189X15618385>.
- American Federation of Teachers, National Council on Measurement in Education, and National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. <https://buros.org/standards-teacher-competence-educational-assessment-students>.
- Arkansas Division of Elementary and Secondary Education. (2023). *Arkansas Initiative for Instructional Materials (AIIM)*. <https://dese.ade.arkansas.gov/Offices/learning-services/curriculum-support/arkansas-initiative-for-instructional-materials-aiim>.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258–267. <https://doi.org/10.3102/0013189X07306523>.
- Booher-Jennings, J. (2005). Below the bubble: “Educational triage” and the Texas accountability system. *American Educational Research Journal*, 42(2), 231–268. <https://doi.org/10.3102/00028312042002231>.
- Brookhart, S. M., & Nitko, A. J. (2019). *Educational assessment of students* (8th ed.). Pearson.
- Buzick, H. M., Casabianca, J. M., & Gholson, M. L. (2023). Personalizing large-scale assessment in practice. *Educational Measurement: Issues and Practice*, 42(2), 5–11. <https://doi.org/10.1111/emip.12551>.
- California Department of Education. (n.d.). *Curriculum frameworks & instructional materials: Information on all subject areas*. <https://www.cde.ca.gov/ci/cr/cf>.
- Chappuis, S., Commodore, C., & Stiggins, R. (2016). *Balanced assessment systems: Leadership, quality, and the role of classroom assessment*. Corwin.
- Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives*, 9(4), 173–206. <https://doi.org/10.1080/15366367.2011.626729>.
- Coladarci, T. (2002). Is it a house ... or a pile of bricks? Important features of a local assessment system. *The Phi Delta Kappan*, 83(10), 772–774.
- Common Core State Standards Initiative. (2021). *Key shifts in mathematics*. <https://www.thecorestandards.org/other-resources/key-shifts-in-mathematics>.
- Conley, D. T., & Darling-Hammond, L. (2013). *Creating systems of assessment for deeper learning*. Stanford Center for Opportunity Policy in Education. [https://edpolicy.stanford.edu/sites/default/files/publications/creating-systems-assessment-deeper-learning\\_0.pdf](https://edpolicy.stanford.edu/sites/default/files/publications/creating-systems-assessment-deeper-learning_0.pdf).
- Council of Chief State School Officers. (2015). *Comprehensive statewide assessment systems: A framework for the role of the state education agency in improving quality and reducing burden*. <https://www.ccsso.org/sites/default/files/2017-12/Comprehensive%20Statewide%20Assessment%20Systems%20-%20A%20Framework%20Final%206-24.pdf>.
- Council of Chief State School Officers. (2020). *State Policy on Educator Preparation: A Framework for Levers to Consider*. <https://ccsso.org/sites/default/files/2022-01/EPP%20Framework.pdf>.
- Council of Chief State School Officers. (2021a). *High quality instructional materials & professional development network case study: Communications & stakeholder engagement*. <https://753a0706.flowpaper.com/CCSSOIMPDCaseStudyCommunications/#page=1>.
- Council of Chief State School Officers. (2021b). *Overview of CCSSO's high quality instructional materials & professional development network: How states can support the adoption and effective use of high-quality standards aligned instructional materials*. <https://753a0706.flowpaper.com/CCSSOIMPDCaseStudyOVERVIEWFINAL/#page=1>.

- Council of Chief State School Officers. (2022a). *Guidance for states on supporting district use of high-quality professional learning providers*. <https://drive.google.com/file/d/1FvJdsUnTeab1apY91z9QM6h0BLr0cg1W/view?usp=sharing>.
- Council of Chief State School Officers. (2022b). *CCSSO's high quality instructional materials & professional development network case study: An IMPD policy roadmap for states*. <https://753a0706.flowpaper.com/CCSSOIMPDPolicyRoadmap/#page=1>.
- Council of Chief State School Officers. (2022c). *High quality instructional materials & professional development network case study: Impact of the CCSSO IMPD network*. <https://753a0706.flowpaper.com/CCSSOIMPDCaseStudyImpact/#page=1>.
- Council of Chief State School Officers. (2022d). *High quality instructional materials & professional development network case study: Signaling & incentivizing quality*. <https://753a0706.flowpaper.com/CCSSOIMPDCaseStudyIncentivizing/#page=1>.
- Council of Chief State School Officers. (n.d.). *CCSSO's high-quality instructional materials and professional development (IMPD) network*. <https://learning.ccsso.org/high-quality-instructional-materials>.
- Dadey, N., Evans, C. M., & Lorie, W. (2023). *Through-year assessment: Ten key considerations*. National Center for the Improvement of Educational Assessment. <https://www.nciea.org/wp-content/uploads/2023/03/Ten-Key-Considerations-Through-Year-Assessment-Report-March2023-F.pdf>.
- Dadey, N., & Gong, B. (2017). *Using interim assessments in place of summative assessments? Consideration of an ESSA option*. Council of Chief State School Officers. [https://ccsso.org/sites/default/files/2017-12/ASR\\_ESSA\\_Interim\\_Considerations-April.pdf](https://ccsso.org/sites/default/files/2017-12/ASR_ESSA_Interim_Considerations-April.pdf).
- Dadey, N., & Gong, B. (2021). *Claims and evidence for through year assessments: Definitions, aims, and use cases*. <https://www.nciea.org/wp-content/uploads/2022/07/Through-Year-Convening-Session-1-Slides.pdf>.
- Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record*, 117(4), 1–26. <https://doi.org/10.1177/016146811511700408>.
- DeLuca, C., Coombs, A., MacGregor, S., & Rasooli, A. (2019). Toward a differential and situated view of assessment literacy: Studying teachers' responses to classroom assessment scenarios. *Frontiers in Education*, 4, 1–10. <https://doi.org/10.3389/feduc.2019.00094>.
- Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record*, 106(6), 1145–1176. <https://doi.org/10.1111/j.1467-9620.2004.00375.x>.
- Doan, S., Kaufman, J. H., Woo, A., Tuma, A. P., Diliberti, M. K., & Lee, S. (2022). *How states are creating conditions for use of high-quality instructional materials in K–12 classrooms: Findings from the 2021 American Instructional Resources Survey*. RAND Corporation. <https://doi.org/10.7249/RR134-13>.
- EdReports. (2022a). *Become a reviewer*. <https://www.edreports.org/process/reviewers>.
- EdReports. (2022b). *Review tools*. <https://www.edreports.org/reports/review-tools>.
- EducationFirst. (n.d.). *Fewer and better local assessments: A toolkit for educators (playbook)*. <https://www.education-first.com/playbook>.
- Evans, C. M. (2022, June 22). *Can a state test inform instruction? The typical, the improbable, and the possibilities in the messy middle*. National Center for the Improvement of Educational Assessment. <https://www.nciea.org/blog/can-a-state-test-inform-instruction>.
- Evans, C. M., & Marion, S. F. (in press). *Instructionally Useful Assessment*. Routledge.
- Evans, C. M., & Thompson, J. (2022a). *Educational assessment 101 for K–12 teachers and leaders (version 2.0)*. National Center for the Improvement of Educational Assessment. <https://www.nciea.org/library/classroom-assessment-learning-modules>.
- Evans, C. M., & Thompson, J. (2022b). *Leader module 9: What is a local assessment system and how can school/district leaders audit their system to ensure it aligns with the features of balanced assessment systems?* National Center for the Improvement of Educational Assessment. <https://docs.google.com/document/d/1uCHOeYIc6fELyEJma68VeF38-j7NvG556WGY01OtSkA/edit>.
- Evans, C. M., & Thompson, J. (2022c). *Teacher module 23: How do you know if you have a high-quality classroom assessment system?* National Center for the Improvement of Educational Assessment. [https://docs.google.com/document/d/17rvK7mZkZAVBpj\\_WPxTweGaFjm0j6oEYTURa7bkvnA/edit](https://docs.google.com/document/d/17rvK7mZkZAVBpj_WPxTweGaFjm0j6oEYTURa7bkvnA/edit).



- 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](https://www.rand.org/pubs/research_reports/RR354.html).
- Firestone, W. A., Camili, G., Yurecko, M., Monfils, L., & Mayrowetz, D. (2000). State standards & opportunity to learn in New Jersey. *Education Policy Analysis Archives*, 8(35). <https://doi.org/10.14507/epaa.v8n35.2000>.
- Fuhrman, S. H., & Elmore, R. F. (Eds.). (2004). *Redesigning accountability systems for education*. Teachers College Press.
- Georgia Partnership for Excellence in Education. (2022). *Assessment inventory project*. <https://gpee.org/programs/testing-inventory-project>.
- Gong, B. (2010). *Using balanced assessment systems to improve student learning and school capacity: An introduction*. Council of Chief State School Officers and Renaissance Learning.
- Gong, B. (2019). *Matching instructional uses with interim assessment designs*. <https://www.nciea.org/wp-content/uploads/2021/11/Gong-Matching-Instructional-Uses-with-Assessment-Designs-100719.pdf>.
- Hawai'i Performance Assessment Task Bank. (2022). *Hawai'i performance assessment development initiative: Performance assessment task bank & administration guidance*. <https://sites.google.com/view/hawaiiassessmentbank/home>.
- Kaufman, J. H., Tosh, K., & Mattox, T. (2020). *Are U.S. teachers using high-quality instructional materials?* RAND Corporation. <https://doi.org/https://doi.org/10.7249/RR2575.11-1>.
- Klinger, D. A., McDivitt, P. R., Howard, B. B., Munoz, M. A., Rogers, W. T., & Wylie, E. C. (2015). *The classroom assessment standards for PreK–12 teachers*. Kindle Direct Press. <https://evaluationstandards.org/classroom>.
- Landl, E., & Lyons, S. (2023). *Choosing the right tests: The District Assessment Procurement Protocol (DAPP)*. EdReports.org, Center for Assessment, and Lyons Assessment Consulting. [https://www.lyonsassessmentconsulting.com/assets/files/District\\_Assessment\\_Procurement\\_Protocol\\_FINAL-1.pdf](https://www.lyonsassessmentconsulting.com/assets/files/District_Assessment_Procurement_Protocol_FINAL-1.pdf).
- Lazarin, M. (2014, October 16). *Testing overload in America's schools*. Center for American Progress. <https://www.americanprogress.org/article/testing-overload-in-americas-schools>.
- Louisiana Department of Education. (n.d.). *Online instructional materials review*. <https://www.louisianabelieves.com/academics/ONLINE-INSTRUCTIONAL-MATERIALS-REVIEWS>.
- Marion, S. F. (2018). The opportunities and challenges of a systems approach to assessment. *Educational Measurement: Issues and Practice*, 37(1), 45–48. <https://doi.org/10.1111/emip.12193>.
- Marion, S. F. (2019a, October 30). *Do interim assessments have a role in balanced systems of assessment?* National Center for the Improvement of Educational Assessment. <https://www.nciea.org/blog/assessment/do-interim-assessments-have-role-balanced-systems-assessment>.
- Marion, S. F. (2019b, March 21). *Five essential features of assessment for learning*. National Center for the Improvement of Educational Assessment. <https://www.nciea.org/blog/innovative-assessment/five-essential-features-assessment-learning>.
- Marion, S., Thompson, J., Evans, C., Martineau, J., & Dadey, N. (2019). *The challenges and opportunities of balanced systems of assessment: A policy brief*. National Center for the Improvement of Educational Assessment. <https://files.eric.ed.gov/fulltext/ED598421.pdf>.
- Martineau, J., Dewsbury-White, K., Roeber, E., Vorenkamp, E., Snead, S., & Flukes, J. (2018). *District assessment system design toolkit*. National Center for the Improvement of Educational Assessment. <https://www.nciea.org/library/district-assessment-system-design-toolkit>.
- Massachusetts Department of Elementary and Secondary Education. (2023). *Mathematics released items*. <http://mcas.pearsonsupport.com/released-items/math>.
- Mathematics Assessment Resource Service. (2011). *Mathematics assessment project: Balanced assessment: Summative assessment tasks for high school*. <https://www.map.mathshell.org/tasks.php?unit=HA07&collection=9>.
- McMillan, J. H. (2021). *Classroom assessment: Principles and practice that enhance student learning and motivation* (7th ed.). Pearson.
- Michigan Assessment Consortium (2016). *Assessment literacy standards*. <https://www.michiganassessmentconsortium.org/assessment-literacy-standards>.
- 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. (2003). *Assessment in support of instruction and learning: Bridging the gap between large-scale and classroom assessment: Workshop report*. The National Academies Press. <https://doi.org/10.17226/10802>.
- 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 Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press. <https://doi.org/10.17226/13165>.
- Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, 28(3), 5–13. <https://doi.org/10.1111/j.1745-3992.2009.00149.x>.
- Polikoff, M. (2021). *Beyond standards: The fragmentation of education governance and the promise of curriculum reform*. Harvard Education Press.
- Shapovalov, Y. A., & Evans, C. M. (2022). *Research synthesis: Developing the educator assessment literacy professional learning screener*. National Center for the Improvement of Educational Assessment. <https://files.eric.ed.gov/fulltext/ED627088.pdf>.
- Shepard, L. A. (2005). *Formative assessment: Caveat emptor*. ETS Invitational Conference, The Future of Assessment: Shaping Teaching and Learning. <https://csaa.wested.org/wp-content/uploads/2019/11/shepard-formative-assessment-caveat-emptor.pdf>.
- Shepard, L. A. (2021). Ambitious teaching and equitable assessment: A vision for prioritizing learning, not testing. *American Educator*, Fall. <https://www.aft.org/ae/fall2021/shepard>.
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, 37(1), 21–34. <https://doi.org/10.1111/emip.12189>.
- Shultz, P. K., & Englert, K. (2021). Cultural validity as foundational to assessment development: An Indigenous example. *Frontiers in Education*, 6. <https://doi.org/10.3389/educ.2021.701973>.
- Silver, D. (2022). A theoretical framework for studying teachers' curriculum supplementation. *Review of Educational Research*, 92(3), 455–489. <https://doi.org/10.3102/00346543211063930>.
- State of Connecticut. (2023). *About: Center for Connecticut Education Research Collaboration*. <https://portal.ct.gov/ccerc/About-Us>.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72, 534–539.
- Stiggins, R. (2006). *Balanced assessment systems: Redefining excellence in assessment*. Educational Testing Service. [https://media.ride.ri.gov/PD/FA/Formative\\_Assessment\\_Module\\_1\\_Lesson\\_4/story\\_content/external\\_files/Stiggins\\_Balanced\\_Assessment\\_Systems%202006.pdf](https://media.ride.ri.gov/PD/FA/Formative_Assessment_Module_1_Lesson_4/story_content/external_files/Stiggins_Balanced_Assessment_Systems%202006.pdf).
- Stiggins, R. (2008). *Assessment manifesto: A call for the development of balanced assessment systems*. ETS Assessment Training Institute. <https://famemichigan.org/wp-content/uploads/2018/06/Stiggins-Assessment-Manifesto-A-Call-for-the-Development-of-Balanced-Assessment-Systems.pdf>.
- Stiggins, R. (2017). *The perfect assessment system*. ASCD.
- Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change*, 10, 211–227. <https://doi.org/10.1007/s10833-009-9105-2>.
- The Leadership Conference on Civil and Human Rights. (2015). *Civil rights groups: "We oppose anti-testing efforts."* <https://civilrights.org/2015/05/05/civil-rights-groups-we-oppose-anti-testing-efforts>.
- The Office of Governor Ned Lamont. (2023, March 22). *Governor Lamont announces plan to launch intensive mathematics tutoring program for students in grades 6–9 to address learning loss from pandemic*. Press release. <https://portal.ct.gov/Office-of-the-Governor/News/Press-Releases/2023/03-2023/Governor-Lamont-Announces-Plans-to-Launch-Intensive-Mathematics-Tutoring-Program>.
- U.S. Department of Education. (2022). *Awards*. <https://oese.ed.gov/offices/office-of-formula-grants/school-support-and-accountability/competitive-grants-for-state-assessments/awards>.
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>.

# Developing, Implementing, and Institutionalizing Complex Educational Innovations: Considerations for Balanced Assessment Systems

Donald J. Peurach, *University of Michigan*

Jennifer Lin Russell, *Vanderbilt University*

## CONTENTS

INTRODUCTION .....	238
CONTEXT AND FRAMING .....	239
The Idea and Challenges of Balanced Assessment Systems .....	239
Reconsidering Development, Implementation, and Institutionalization.....	240
Our Analytic Framework and Approach.....	241
SEEING AND CRAFTING AT THE POLICY LEVEL.....	244
Resource-Forward Innovation and Improvement .....	245
Practice-Forward Innovation and Improvement .....	247
Empowerment-Forward Innovation and Improvement.....	249
Crafting Coherence.....	252
SEEING AND CRAFTING AT THE LOCAL LEVEL.....	254
School Systems.....	255
Education Systems.....	257
Learning Systems.....	259
Crafting Coherence.....	261
LEARNING WHILE LEADING.....	263
The Learning Imperative.....	263
The Learning Lattice .....	264
A Lattice of Learning Opportunities for State Leaders.....	265
A Lattice of Learning Opportunities for Local Leaders.....	266
KEY TAKEAWAYS .....	267
REFERENCES .....	267

## INTRODUCTION

Beginning in the 1980s, increasing recognition of the complex, interdependent problems underlying persistently low quality and inequitable educational outcomes in the United States has given rise to increasingly complex innovations that aim to address multiple problems simultaneously, in interaction (Peurach, 2011). Examples of such innovations include “whole school” / school-wide reforms, comprehensive school reform programs, school turnaround models, charter school networks, and networked improvement communities.

Developing, implementing, and institutionalizing complex innovations is no simple matter. Doing so involves coordinated efforts among interdependent actors distributed among multiple organizations, each with their own agendas and constituencies; in variable authority and influence relationships; with different stocks and flows of resources; in diffuse, dynamic educational environments; and over long periods of time. Yet any such efforts rest atop a fundamental problem: The very complexity of many of these innovations often makes them difficult to perceive and to understand, never mind to develop, implement, and institutionalize.

Balanced assessment systems are a case in point. If nothing else, this volume is a representation of balanced assessment systems as a complex innovation that aims to address multiple problems simultaneously, in interaction. For example, Chapter 7 in this volume, “State Practices and Balanced Assessment Systems,” theorizes and seeks to guide interdependent state-level activities needed to establish conditions that would support productive engagement with balanced assessment systems in districts and schools. Chapter 6 in this volume, “District and School Practices and Assessments to Support a Learning-Centered Vision,” theorizes and seeks to guide districts and schools in supporting productive engagement in classrooms. Chapter 3 in this volume, “Human Learning and Development: Theoretical Perspectives to Inform Assessment Systems,” elaborates a novel, ambitious, and evolving theory of human learning and development that states, districts, schools, and classrooms should heed in supporting the development of students.

Yet, as argued in Chapter 2 of this volume, “The Struggle to Implement Balanced Assessment Systems: Explanations and Opportunities,” the more than 20-year history of balanced assessment systems to date is largely one of low-level, problematic, and non-engagement, owing, in part, to the complexity of the underlying ideas. As a first recommendation for moving forward in Chapter 2 of this volume, Polikoff and Hutt argue that “achieving balance must be made both more understandable and feasible for educators and local and state policy makers. The criteria underlying balanced assessment systems are laudable, but the ideas are too complex for widespread comprehension and implementation in the current highly decentralized, capacity-poor education systems” (p. 43).

Indeed, as argued in Chapter 1 of this volume, “Reimagining Balanced Assessment Systems: An Introduction,” the idea of balanced assessment systems has been distorted (and sometimes corrupted) as it has been taken up not only by state and local actors within the formal educational governance structure but also by market actors. This volume as a whole is premised on the concern that this distortion and corruption is

sufficient to warrant a new and clearer articulation of the idea (see Chapter 1 of this volume).

The purposes of this chapter, thus, are (a) to establish a general framework for understanding complex educational innovations and (b) to model the use of this framework for understanding the fundamental ideas underlying balanced assessment systems. Given the practical ambitions of this volume, one instrumental aim is to guide the integration of balanced assessment systems into ongoing reform efforts at the state and local levels that aspire to advance educational quality and equity.

This chapter is structured in four sections. The first establishes context, introduces our analytic framework, and sets out our plans for further developing it. The second and third sections develop (and model the use of) the framework at the state and local levels. The fourth section discusses considerations for the learning and engagement of state and local leaders.

We conclude with what we see as the key takeaway from this chapter and from the volume: The adult and organizational learning demands of balanced assessment systems are every bit as novel and ambitious as the goals for student learning, and those learning demands require commensurate attention in developing, implementing, and institutionalizing balanced assessment systems.

## CONTEXT AND FRAMING

Our premise is that, while complex, neither the ideas underlying balanced assessment systems nor the broader environments in which they operate are unknowable. The matter at hand is developing a schema and approach for learning about them.

### The Idea and Challenges of Balanced Assessment Systems

Again, we take the fundamental task for state and district leaders to be building shared meaning and understanding of the idea of balanced assessment systems as a prerequisite to supporting their development, implementation, and institutionalization. If that is their fundamental task, it promises to be a challenging one, beginning with the complexity of the idea itself. Indeed, the idea of balanced assessment systems (as it was first introduced, as it has evolved, and as represented across this volume) can be understood as having three core dimensions: *mechanisms*, *practical ambitions*, and a *theory of action*.

- The *mechanisms* of balanced assessments systems are resources: integrated assessments designed to serve different purposes among different actors at the classroom, school, district, and state levels, all anchored in theories of student learning that center the development of the whole child (cognitive, social, and emotional) and the communities in which children live.
- The *practical ambitions* of balanced assessment systems are to coordinate the work of organizing, managing, and improving instruction from the classroom level to the state level, as enacted by teachers, local leaders, and state leaders, to advance quality and equity in students' educational opportunities, experiences, and outcomes. Central to this work is advancing ambitious teaching commensurate

- with theories of student learning that center the development of the whole child in community.
- The *theory of action* is that assessment at all levels will provide teachers, local leaders, and state leaders with the evidence that they need both to (a) work within existing systems to make incremental adaptations to students' learning opportunities and experiences and (b) bring diverse perspectives to bear on interrogating, disrupting, and reforming existing systems to support cognitively rigorous, socially and culturally relevant, and personalized learning for all students.

Another challenge lies in the complex U.S. public education enterprise into which state and local leaders are to introduce the idea of balanced assessment systems. This enterprise includes federal and state governments, with educational responsibilities distributed among levels, branches, and agencies. It includes a national-level market that has long served as a primary source of material, human, and knowledge resources. It includes professional associations, interest groups, advocacy organizations, philanthropies, think tanks, and research institutes that seek to inform and influence political, policy, and social agendas. And it includes public school districts with educational responsibility distributed among central offices, schools, and classrooms, themselves remarkably variable in form and governance.

Yet another challenge is that the U.S. public education enterprise has been in the throes of active reform since the mid-20th century in pursuit of the very goals of balanced assessment systems, with no signs of stopping. Indeed, the idea of balanced assessment systems is one among many policy-level initiatives, past and present, pressing local districts and schools to organize, manage, and improve instruction in ways that advance educational quality and equity. Some of these policy-level initiatives have had shared aspirations for aligned coherent systems. Many others have not. Rather, the conventional narrative is that policy-level fragmentation, incoherence, and turbulence have fueled faddism in educational innovation and improvement, with pendulums swinging back and forth between competing ambitions and with waves of reform washing in and out. With that, policy-level fragmentation, incoherence, and turbulence are recreated within districts and schools.<sup>1</sup>

### **Reconsidering Development, Implementation, and Institutionalization**

For state and local leaders advancing balanced assessment systems, the matter is not to steward the type of sequential "development, implementation, and institutionalization" process commonly associated with large scale innovations. Moreover, it is

---

<sup>1</sup> Regarding what we describe here and throughout this chapter as the "conventional narrative" about the relationship between policy-level and local-level reform activity: See Smith and O'Day (1990) and Fuhrman (1993) for early and seminal theory and analysis of coherent education policy, including the problems to be addressed through policy coherence. See Bryk et al. (1999), Cohen and Spillane (1992), Hess (1998), Payne (2008), and Powell et al. (1985) for analyses of fragmentation, incoherence, turbulence, and their consequences.



unlikely that balanced assessments systems will cut through the complexity of U.S. public education en route from design to use, nor that they will catalyze coherence throughout. Rather, the risk is that balanced assessment systems will become entangled in the complexity of U.S. public education and fuel further fragmentation, incoherence, and turbulence.

Again, the first-order matter is to build shared meaning and understanding of *the very idea* of balanced assessment systems, so that multiple organizations, interests, and actors across this vast educational landscape with different histories, allegiances, and capabilities recognize and value the place and role of balanced assessment systems in advancing educational quality and equity. From a practical perspective, building shared meaning and understanding will benefit from state and local leaders collaborating on three tasks:

- *Seeing systems*: Analyzing the complex policy and local contexts into which they will be introducing the idea of balanced assessment systems to discern different frames that will shape how educators make sense of the idea.
- *Crafting coherence*: Developing shared understandings of the place and role of balanced assessment systems among other ongoing policy-level and local-level initiatives aiming to advance educational quality and equity.
- *Learning while leading*: Developing opportunities for their own collegial learning, both to share and leverage successes and to work through the inevitable false starts, variable uptake, and difficult-to-discern progress endemic to building shared meaning and understanding of such a complex idea in such a complex context.<sup>2</sup>

### **Our Analytic Framework and Approach**

We continue, then, by developing a general analytic framework for considering the development, implementation, and institutionalization of complex, systemic innovations in interdependent macro-level and local-level education contexts. As we do, we consider the different potential uses of this framework by state and district leaders in building shared meaning and understanding of balanced assessment systems. To develop our analytic framework, we synthesized three interdependent lines of our scholarship and research:

- scholarship on the co-evolution of policy and local contexts in organizing, managing, and improving instruction;
- empirical research on building and rebuilding systems to organize, manage, and improve instruction; and

---

<sup>2</sup> Our notion of “seeing systems” draws from Bryk et al. (2015). Our notion of “crafting coherence” draws from Honig and Hatch (2004) and Russell and Bray (2013).

- improvement research aimed at developing and leveraging capabilities for collaborative, continuous learning and improvement in states, networks, districts, and schools.<sup>3</sup>

Across these three lines of research and scholarship, we have identified order and structure in macro-level policy contexts, local-level systems, and their relationships that may not be immediately apparent at particular moments in time. We have also identified how this order and structure has accumulated historically and how it continues to accumulate. We have then used these insights to construct a counter-narrative to the conventional characterization of policy-level and local-level fragmentation, incoherence, and turbulence in the U.S. public education enterprise.

As summarized in Table 8-1, our analytic framework includes three primary components. The first component is leading policy logics that have been accumulating at the state and national policy level since the mid-1900s: what we call *resource-forward*, *practice-forward*, and *empowerment-forward* innovation and improvement. The second component associates each policy logic with a particular category of organizational legitimacy pressing on local-level efforts to organize, manage, and improve instruction: what we call *structural/procedural*, *technical*, and *moral* legitimacy.<sup>4</sup> The third component charts the co-evolution of local districts in interaction with these policy logics and associated legitimacies: what we describe as a progression from *school* systems to *education* systems to *learning* systems.

The three components of Table 8-1 summarize our historical analysis of the accumulation and co-evolution of (a) national-level policy logics, (b) policy presses on local districts, and (c) capabilities in local districts. It addresses the period following World War II (and the onset of increased federal engagement in public education) to the present. The table can be used as a framework for analyzing the development of logics, presses, and capabilities in individual states and districts and, with that, the schema that shape how states and districts perceive and understand the multi-dimensional idea of balanced assessment systems.

---

<sup>3</sup> These three lines of scholarship and research include systematic reviews and analytic essays that index and integrate multiple literatures on education innovation, policy, reform, and improvement at the policy and local levels, both historical and contemporary. As such, the general framework that we develop here is a “synthesis of syntheses” that is more comprehensive than any of our prior work. In developing this general framework, page limits preclude fully reconstructing our prior reviews and analyses in the context of this chapter. Instead, we include liberal references and notes throughout, trusting that readers who are curious or critical will review our earlier work to suit their needs and interests. By way of overview, for scholarship on the co-evolution of policy and local contexts in organizing, managing, and improving instruction, see Cohen et al. (2018); Peurach et al. (2019b, 2022a); and Spillane et al. (2019a). For empirical research on building and rebuilding systems, see Datnow et al. (2022); Peurach et al. (2019a, 2019c); Russell and Bray (2013); and Spillane et al. (2019b, 2022). For improvement research aimed at developing capabilities for collaborative, continuous learning and improvement, see Peurach et al. (2018, 2022b) and Russell et al. (2015, 2017, 2019, 2020).

<sup>4</sup> Our conceptions of legitimacy are adapted from Spillane et al. (2022) and Suchman (1995).

**TABLE 8-1 Analytic Framework: The Co-Evolution of Policy Logics and Local Districts (post–World War II to present)**

Policy Logics	Policy Press on Local Districts	Local Capabilities	Balanced Assessment Systems
<p><i>Resource-Forward Innovation and Improvement:</i></p> <ul style="list-style-type: none"> <li>• Advance quality and equity via the production of more and better educational resources distributed more equitably among local districts</li> </ul>	<p><i>Structural/Procedural Legitimacy:</i></p> <ul style="list-style-type: none"> <li>• Maintaining good standing by publicly signaling commitments to advancing educational quality and equity and (b) complying with organizational and administrative requirements</li> </ul>	<p><i>School Systems:</i></p> <ul style="list-style-type: none"> <li>• Provide access to instruction by sorting students into schools, grade levels, and classrooms; resourcing them with materials; and delegating responsibility to teachers</li> </ul>	<p><i>Mechanisms:</i></p> <ul style="list-style-type: none"> <li>• Integrated assessments designed to serve different purposes among different actors at the classroom, local, district, and state levels to support the development of the whole child in community</li> </ul>
<p><i>Practice-Forward Innovation and Improvement:</i></p> <ul style="list-style-type: none"> <li>• Advance quality and equity by improving instruction and the school and district contexts in which it is situated</li> </ul>	<p><i>Technical Legitimacy:</i></p> <ul style="list-style-type: none"> <li>• Maintaining good standing and signaling appropriate engagement by generating evidence of improving quality and reducing disparities in student outcomes</li> </ul>	<p><i>Education Systems:</i></p> <ul style="list-style-type: none"> <li>• Districts, schools, and teachers collaborate to organize, manage, and improve instruction to improve quality and reduce disparities</li> </ul>	<p><i>Practical Ambitions:</i></p> <ul style="list-style-type: none"> <li>• Coordinate the organization, management, and improvement of instruction from the state level to the classroom level to advance ambitious teaching</li> </ul>
<p><i>Empowerment-Forward Innovation and Improvement:</i></p> <ul style="list-style-type: none"> <li>• Advance quality and equity by developing and mobilizing local agency and capability, broadly and inclusively, to identify and address local educational ambitions, needs, and problems</li> </ul>	<p><i>Moral Legitimacy:</i></p> <ul style="list-style-type: none"> <li>• Maintaining good standing and signaling appropriate engagement by supporting inclusion and participation among people, groups, and communities whose perspectives and priorities have historically been marginalized</li> </ul>	<p><i>Learning Systems:</i></p> <ul style="list-style-type: none"> <li>• Engage diverse stakeholders (professional, family, and community) in developing understandings, knowledge, and values needed to identify and address local educational ambitions, needs, and problems</li> </ul>	<p><i>Theory of Action:</i></p> <ul style="list-style-type: none"> <li>• Comprehensive, coherent, and continuous assessment will bring diverse perspectives to bear on working within systems and interrogating, disrupting, and reforming systems to advance quality and equity</li> </ul>

We discuss Table 8-1 in detail throughout this chapter. The table supports three lines of analysis:

- As a *synthesis of research*, the first three columns of Table 8-1 can be read from top to bottom as an historical taxonomy of policy logics, presses, and capabilities that are accumulating as actors at the policy and local levels gain increasing understandings of what more is needed to define and advance ambitions for educational quality and equity.
- In practice, these logics, presses, and capabilities will be variably developed in individual states and local districts (e.g., institutionalized, developing, or emerging). With that in mind, the first three columns of Table 8-1 can be read as a *developmental sequence* that can be used to analyze progress in individual states and local districts.
- The levels at which understandings and capabilities are developed, in turn, can be considered as *schema* that enable and constrain ways that state and local actors perceive and understand new ideas.

We also include a fourth component in Table 8-1 framing how state and local actors are apt to perceive and understand the multi-dimensional idea of balanced assessment systems, depending on their level of development. The more developed the policy logics in a state (or the local capabilities in a district), the greater the potential to fully perceive and understand the idea of balanced assessment systems; the less developed, the greater the risk that perception and understanding will be capped at the current level of development. For example, a state that has evolved in ways that also embrace a practice-forward logic (or a local district that has evolved as an education system) is more likely to perceive and to understand the practical ambitions of balanced assessment systems and to be positioned to understand more fully the theory of action. By contrast, a state that operates largely within a resource-forward logic (or a local district that has developed only as a school system) is less likely to perceive and to understand the practical ambitions and theory of action of balanced assessment systems.

By helping them to see order and structure in policy-level and local-level contexts, our analytic framework is a potential resource for state and local leaders in developing shared understandings of the place and role of balanced assessment systems among other policy-level and local-level initiatives, past and present, aiming to advance educational quality and equity. It also is a potential resource for state and local leaders in learning while leading, as they engage colleagues, constituents, stakeholders, and (importantly) each other in building shared meaning and understandings of balanced assessment systems.

### SEEING AND CRAFTING AT THE POLICY LEVEL

The aim of seeing systems and crafting coherence at the state level is to develop shared understandings of the place and role of balanced assessment systems among other ongoing, policy-level initiatives seeking to advance educational quality and equity. Doing so requires gaining perspective on the national education policy context in which states operate. Indeed, as a policy initiative, the idea of balanced assess-

ment systems did not emerge and gain currency within individual states operating in isolation. Rather, it emerged and gained currency through the collaborative efforts of coalitions of non-public and public actors operating at the national level, across states.

As described above, this national policy context is less of a formal *system*: a collection of organizations designed and structured to work in interaction, with shared purpose and toward a common goal. Rather, it is more of a complex *national education ecology*: a sprawling organizational field in which diverse organizations with their own ambitions, interests, and agendas for public education interact across states with varying degrees of mutual awareness, cooperation, and competition; with varying means of influencing local education contexts; and with the federal government as but one player.<sup>5</sup>

Again, the conventional narrative is that this national education ecology is rife with fragmentation, incoherence, and turbulence. This conventional narrative, in turn, is likely to be the cognitive frame through which many perceive and understand balanced assessment systems: that is, as another set of initiatives among many being advanced within and beyond governments; as amplifying demands and disorder; and, thus, as much problem as solution.

Yet, as represented in Table 8-1, we have identified policy logics that have been accumulating at the national level since the mid-1900s, one atop the other, and that bring order and structure to the national education ecology: the logics of *resource-forward*, *practice-forward*, and *empowerment-forward* innovation and improvement. These logics, in turn, create a press on local districts to maintain their *structural/procedural*, *technical*, and *moral* legitimacy.

These policy logics are lines of reasoning that associate policy actions and activities with policy goals and objectives: for example, if we enact policy X (e.g., universal access to preschool), then we will accomplish goal Y (e.g., more equitable student outcomes in K–12 schools). These policy logics structure discourse and debate about policy issues. With increasing consensus, they begin to function as shared assumptions—sometimes explicit, sometimes tacit—about relationships between policy and outcomes, and they incentivize and legitimize local approaches to innovation and improvement.

Our central line of argument is that, by seeing and understanding the national policy context as an ecology structured by these policy logics, state and local leaders will be better positioned to construct a counter-narrative that positions balanced assessment systems squarely within leading lines of reasoning and action aimed at advancing educational quality and equity.

### **Resource-Forward Innovation and Improvement**

The theory of action underlying *resource-forward innovation and improvement* is that students' educational opportunities, experiences, and outcomes can be improved (and disparities among them reduced) by the production of more and better educational resources that are distributed more equitably among local public school districts. By *educational* resources, we mean those integral to the day-to-day work of teaching and learning in classroom instruction, including material resources (e.g., curricula, text-

---

<sup>5</sup> For distinctions between the concepts of education systems and ecologies, see Datnow et al. (2022) and Fuller and Kim (2022).



books, and assessments); knowledge resources (e.g., research, instructional models, and pedagogical routines); and human resources (e.g., teachers, intervention specialists, and paraprofessionals). The resource-forward logic aligns with the idea of assessment resources as the mechanisms of balanced assessment systems.

In our prior research, we dated the logic of resource-forward innovation and improvement to the mid-20th century, with the onset of increasing federal engagement in public education (Peurach et al., 2022a). Beginning in the 1950s, national policy dynamics were focused squarely on advancing educational access and equity: issues driven to the center of the agenda by the civil rights movement, the women’s rights movement, the disability rights movement, and the war on poverty. By 1982, the result was the development of formidable federal and state legal, policy, and financial infrastructures that ensured universal access to public schooling.

Over this period, while quality and equity played out alongside access and equity as national policy priorities, the result was not the development of commensurate, policy-level *educational infrastructure* ensuring quality and equity in students’ education once in schools (Cohen et al., 2014; Peurach et al., 2019b, 2022a). Key components of such an infrastructure would have included social and political consensus on the means and ends of instruction, along with coordinated instructional models, curricula, materials, assessments, and teacher development for pursuing those means and ends. The absence of such an infrastructure owed much to disagreements in the national education ecology in defining and pursuing educational quality and equity, as well as to deep distrust in central government that limited federal and state efforts to address such matters.

Instead, ambitions for advancing quality and equity were taken up in an educational resource market that supported the exchange among non-governmental organizations (on the supply side) and districts and schools (on the demand side) of the materials, methods, programs, and people needed to constitute, enact, and improve classroom instruction.<sup>6</sup> Examples of these non-governmental organizations include commercial publishers, service providers, non-profit organizations, professional associations, and university research centers and projects. Dependence on the educational resource market owed much to comparative trust in entrepreneurship and free markets as drivers of social progress, as well as customary deference to local control in defining and pursuing educational quality and equity.

The educational resource market dates to the colonial era, when commercial publishers emerged as the primary suppliers of curriculum resources for newly emerging public schools. The educational resource market began developing further (and rapidly) in the second half of the 20th century, in part due to increased federal investment in producing and disseminating more and better resources to support instruction for the more (and more diverse) students gaining access to public schools.

On the demand side, this included federal block grants, formula grants, and categorical grants to states, districts, and schools that provided supplemental and discretionary funding aimed at advancing educational quality and equity, including federal policies that provided funding to support vocational education students, education-

---

<sup>6</sup> For more on the emergence of the educational resource market and its association with the federally funded evidence infrastructure, see Cohen and Mehta (2017); Peurach et al. (2018, 2019b, 2022a); and Rowan (2002).

ally disadvantaged students, and special education students.<sup>7</sup> On the supply side, the expansion included federal grants and contracts aimed at fueling entrepreneurship and innovation among non-governmental and quasi-governmental organizations in order to provide resources for instruction and instructional improvement within policy-prioritized niches. Spanning both the demand and supply sides were federal efforts to develop an “evidence infrastructure” to support the production and dissemination of basic and applied research as inputs both to local practice and to the production of educational resources.

There were few barriers to entry on the supply side of the market, thus enabling the development of what, by the 1990s, would become a multi-billion dollar “school improvement industry” in which for-profit firms, non-profit organizations, and membership organizations provided resources for instruction and instructional improvement to local districts (Rowan, 2002). Even so, on the demand side, federal and state policy had not evolved by the early 1990s to include means of holding districts and schools accountable for actually using new resources to advance educational quality and equity. Instead, federal and state oversight of the use of new resources focused chiefly on the creation of new structures and programs in local districts to serve different categories of students (e.g., special education, Title I, second language, vision and hearing impaired, and gifted and talented) and on demonstrating compliance with associated rules, regulations, and requirements in administering those programs.

Absent accountability for advancing quality and equity, the policy-level press on local districts was to maintain their *structural/procedural legitimacy*: maintaining their good standing both by (a) adopting resources and initiating programs that publicly signaled commitments to advancing educational quality and equity and (b) complying with organizational and administrative requirements of federal and state funding streams.

### **Practice-Forward Innovation and Improvement**

The theory of action underlying *practice-forward innovation and improvement* is that students’ educational opportunities, experiences, and outcomes can be improved (and disparities reduced) by improving instructional practice and the school and district contexts in which it is situated. The practice-forward logic aligns with the idea in balanced assessment systems of coordinating the organization, management, and improvement of instruction in states, districts, schools, and classrooms as essential for advancing educational quality and equity.

In our prior research, we associated the onset of practice-forward innovation and improvement with two loci of activity in the national education ecology. The first was seminal organizational and policy research that, beginning in the 1960s and carrying into the 1980s, provided increasing transparency in the operations, outcomes, and improvement of local public school districts. This included research that evidenced problems and challenges, including:

---

<sup>7</sup> Seminal federal policies include the Vocational Education Act of 1963, the Elementary and Secondary Education Act of 1965, and the Education for All Handicapped Children Act of 1975.

- ways in which local districts maintained structural/procedural legitimacy without improving instruction, its organization, or its management;
- ways in which maintaining structural/procedural legitimacy supported within-school segregation of students into academic tracks and categorical programs, with one consequence being inequities in students' educational opportunities, experiences, and outcomes; and
- a formidable and persistent achievement gap between White and Black students, despite (and, in some cases, because of) the sustained policy focus on resource-forward innovation and improvement.<sup>8</sup>

This also included research that evidenced potential and possibilities, including research on effective schools in which educational expectations, instructional opportunities, leadership, climate, and home-school relationships were coordinated in ways that supported the academic success of historically marginalized students.<sup>9</sup> This research was instrumental in effecting a shift toward “whole school” reform, with entire schools (and, later, districts) as the units of improvement.

The second locus of activity was the onset of an “excellence and equity” movement in the 1980s and early 1990s. The movement was catalyzed by the 1983 publication of *A Nation at Risk: The Imperative for Educational Reform* and energized by the historic 1989 Charlottesville Education Summit, which drove educational quality and equity to the center of the national policy agenda (National Commission on Excellence in Education, 1983; Vinovskis, 1999). The movement evolved further in the early 1990s with the introduction of *systemic reform* as a policy logic that sought to bring a stronger, more coherent instructional focus to the national educational ecology, with coordinated, state-level content standards, performance standards, and accountability assessments both (a) driving alignment through the educational resource market and (b) motivating and guiding school-wide and district-wide improvement (Fuhrman, 1993; Smith & O'Day, 1990). And the movement emerged in interaction with new calls and ideas for moving beyond didactic instruction and basic skills to ambitious teaching and learning for all students (Cohen et al., 1993).

From the mid-1990s to the present, the focus on excellence and equity has been taken up in a litany of federal, state, and national policies, initiatives, and movements.<sup>10</sup> These policies, initiatives, and movements are by no means a coherent, stable policy-level educational infrastructure, and they have certainly introduced problems of their own (as discussed in the next section). Even so, many of these policies, initiatives, and movements have pursued distinct points of leverage that, collectively, have asserted a comprehensive, sustained press on local districts to advance educational quality and equity by organizing, managing, and improving instructional practice in ways that

---

<sup>8</sup> For example, on the disconnect between structural/procedural legitimacy and instructional practice, see Meyer and Rowan (1978). On the adverse effects of tracking and categorical programs, see Allington and Johnston (1989) and Oakes (1985). On the achievement gap and its relationship to resources, see Coleman et al. (1966) and Jencks et al. (1972). On resources and their use, see Cohen et al. (2003).

<sup>9</sup> On effective schools and districts, see Edmonds (1979) and Purkey and Smith (1983, 1985).

<sup>10</sup> Key examples include the Improving America's Schools Act of 1995; the No Child Left Behind Act of 2001; the Common Core State Standards in 2010; the Next Generation Science Standards in 2013; and the Every Student Succeeds Act of 2016.

they had not historically (Peurach et al., 2019b). Leading movements and points of leverage include:

- *Systems thinking* that takes entire schools and districts as the units of improvement and that aims for coherent organizational support for classroom instruction.
- *Standards and accountability* aimed at (a) raising expectations and building consensus around ambitions for student learning, instructional practice, and leadership practice and (b) motivating improvement through incentives and sanctions tied to assessments and evaluations.
- *Markets and choice* in and among public, charter, and other schools aimed at stimulating educational entrepreneurship and innovation that is responsive to the educational values and aspirations of students and families, especially students and families challenged to exercise political and social influence.
- *Data and evidence* aimed at (a) advancing disciplined, data-driven, evidence-informed analysis, planning, and evaluation in local districts and (b) incorporating evidence-based/evidence-proven resources and practices into those efforts.
- *Autonomy and professionalism* aimed at (a) preserving local authority over substantive educational matters and (b) developing teachers' and leaders' knowledge, capabilities, and values as key levers for advancing educational quality and equity (Peurach et al., 2022a).

From the mid-1990s to the present, the focus on excellence and equity has also been taken up in the educational resource market as a source of practical guidance and support, fueled by continuing federal and philanthropic investment on both the supply and demand sides of the market (Hodge et al., 2019). This includes formidable federal investment in the development and adoption of research-based/research-validated resources and programs (Peurach et al., 2018). This federal investment, in turn, has sustained the press on local districts to develop and maintain structural/procedural legitimacy.

At the same time, this sustained, three-decade-long policy focus on improving instructional practice and its school/district contexts introduced a new press on local districts to develop and maintain their *technical legitimacy*: maintaining good standing and signaling appropriate engagement by actually generating evidence of improving quality (and reducing disparities) in student outcomes. The press for technical legitimacy is strongest in academic content areas that are the primary focus of state standards and accountability policies: English language arts, mathematics, and, more recently, science.

### **Empowerment-Forward Innovation and Improvement**

The theory of action underlying *empowerment-forward innovation and improvement* is that educational quality can be improved (and disparities reduced) by developing and mobilizing local agency and capability, broadly and inclusively, to identify and address local educational ambitions, needs, and problems. The empowerment-forward logic aligns with the idea of using balanced assessment systems not only to work within existing systems but also to bringing diverse perspectives to bear on interrogating, disrupting, and reforming existing systems to advance educational quality and equity.

To be sure, local control has been a pillar of U.S. public education since its inception, as has dependence on the educational resource market as a mechanism for advancing educational quality and equity. Yet empowerment-forward innovation and improvement problematizes the tradition of local control and calls into question an exclusive dependence on the educational resource market. As a policy logic, empowerment-forward innovation and improvement is anchored in three core premises:

- weaknesses in local agency and local capability for addressing complex educational ambitions, needs, and problems are pervasive;
- local agency and local capability are distributed inequitably and in ways that disenfranchise poor and minoritized communities; and
- policy-level ambitions for advancing educational quality and equity require a commensurate, policy-level focus on redressing variability and inequities in local agency and local capability.<sup>11</sup>

In our prior research, we dated the emergence of empowerment-forward innovation and improvement to the 2000s, in association with three loci of activity in the national education ecology. The first is the rise of critical perspectives on practice-forward innovation and improvement as it had emerged and developed to that point.<sup>12</sup> Of these perspectives, there is no shortage, including concerns with:

- the narrowing of educational purpose to students' academic (versus holistic) development; of academic focus to state-assessed contents areas; and of instruction to test preparation;
- the increasing influence of the federal and state governments, policy elites, and resource providers over local educational matters;
- the emergence of a "failing schools" narrative associating evidence of persistent underperformance with students, schools, and communities of color;
- challenges faced by parents and caregivers of historically marginalized students in exercising choice among charter, public, and other schools; and
- recognition of the systemic disempowerment and harm experienced by minoritized communities and, with that, renewed calls for equal voice and participation in defining and advancing quality and equity in their schools.

The second locus of activity is research on local efforts to organize, manage, and improve instruction to advance educational quality and equity and, with that, to establish and maintain technical legitimacy. This research suggested that demands of practice-forward innovation and improvement often exceeded local capabilities for self-improvement, especially in districts and schools that had long struggled to support the

---

<sup>11</sup> Our conception of empowerment is consistent with Richard Elmore's principle of "reciprocity of accountability for capacity" (2002, p. 5).

<sup>12</sup> For critical perspectives on the logic and enactment of policies that we associate with the practice-forward logic, see Au (2010); Burch (2009); Ishimaru et al. (2019); Reckhow (2012); Spillane and Sun (2020); and Wilson and Horsford (2013).



academic success of historically marginalized students.<sup>13</sup> At the same time, this research also began to identify new ways to organize collaborative learning as distributed in and across districts to support the development of such capabilities.

The third locus of activity is the rise of two movements seeking to develop identity, galvanize support, and shape the agenda at the policy level: a rejuvenated racial and social justice movement and a new “improvement movement.” Both movements share an internal logic: take variability and inequity in educational opportunities, experiences, and outcomes as the fundamental problem; interrogate the systems that produce that variability; and intervene on those systems either incrementally or comprehensively to advance educational quality and equity.

- With roots stretching back to access-focused advocacy in the 1950s, the racial and social justice movement has, as a primary focus, identifying and rectifying systemic inequities both in the national education ecology and in local public school districts. Chief among these inequities is the marginalization and exclusion of people of color in defining and advancing educational quality and equity.
- With roots stretching back to the advent of action research in the 1940s, the “improvement movement” has, as a primary focus, advancing disciplined approaches to collaborative, continuous improvement. These approaches have educational professionals, community members, researchers, and other stakeholders working together to understand and address problems of quality and equity in local educational contexts.<sup>14</sup>

These two movements are being advanced by different organizations and initiatives with distinct points of leverage that, together, are pressing for a policy-level response to variability and inequities in local agency and local capability. For example, these two movements have been championed by such organizations as the Civil Rights Project, the Learning Policy Institute, the Carnegie Foundation for the Advancement of Teaching, and the Bill and Melinda Gates Foundation. These two movements are fueled by active domains of research and scholarship in the education sector (e.g., socio-cultural learning theory, solidarity-driven co-design, improvement science, and design-based implementation research) and beyond (e.g., the advancement of critical and pragmatic approaches to knowledge production). This research and scholarship includes fundamental reconsiderations of human learning and development (National Academies of Sciences, Engineering, and Medicine, 2018; National Research Council, 2000) and of the production of knowledge (Medin et al., 2014). And these two movements have, again, been taken up in the educational resource market as a source of practical guidance and support.

While they have not yet gained solid footing in federal and state policy (and, in some cases, face active opposition), these two movements (especially the racial and social justice movement) have been instrumental in asserting a new press on local dis-

---

<sup>13</sup> For research on the practice and knowledge demands on large-scale instructional improvement, see Cohen and Ball (1999); Cohen et al. (2014); Glazer and Peurach (2015); Peurach and Glazer (2012); Peurach et al. (2016); and Spillane and Thompson (1997).

<sup>14</sup> For an analysis of the relationship between the racial and social justice movement and the improvement movement in the contemporary policy context, see Peurach et al. (2022b).

tricts to establish and maintain their *moral legitimacy*: maintaining good standing and signaling appropriate engagement by cultivating agency, capability, and participation for local innovation and improvement, especially among people, groups, and communities whose perspectives and priorities have been historically marginalized.

### Crafting Coherence

There is plenty in the preceding analyses that plays directly to the conventional narrative of the national education ecology as rife with fragmentation, incoherence, and turbulence. With that comes a risk that the idea of balanced assessment systems will be seen as more of the same.

For state leaders charged with supporting the idea of balanced assessment systems, there is advantage in mitigating that risk by collaborating with colleagues to ensure operational alignment among policy-level initiatives seeking to advance educational quality and equity: for example, standards-and-accountability regimes, teacher and leader evaluation policies, public-facing data dashboards, and curriculum frameworks and instructional guidance. In some states, such work may be already underway.

Even so, one problem is that state-level efforts to achieve operational alignment among policy initiatives contends with other sources of influence from the national education ecology in shaping understandings of balanced assessment systems, including the marketing campaigns of commercial assessment providers and advocacy campaigns from national testing consortia, professional associations, and academic associations. Another problem is that such efforts often come up short (e.g., Coburn et al., 2016; Polikoff, 2012a, 2012b, 2015). Yet another problem is that, even if achieved, operational alignment at the policy level is unlikely to mitigate the risk that the idea of balanced assessment systems will be seen as more of the same. As argued by Emily Hodge and Elizabeth Stosich in a study of local engagement with the Common Core State Standards:

Even policies that appear coordinated may not be experienced as such by educators. This presents a challenge as educators are unlikely to devote the necessary attention and resources to implementing policies that they view as disconnected or contradictory, which can result in limited attention to particular policy goals and little, if any, change to practice.... Successful policy implementation requires both *aligning*, or “lining up” policy expectations, resources, and rewards/sanctions, and creating a sense of *coherence*, or a perception that policies are consistent and comprehensible to those who experience them. (Hodge & Stosich, 2022, p. 544)

For state leaders, crafting coherence centers the work of framing, narrating, and sensegiving: actively shaping how policy-level and local-level actors see, understand, and value the place and role of balanced assessment systems in the national education ecology, especially among other initiatives seeking to advance educational quality and equity (Weick, 1995; Weick et al., 2005). To the extent that the analytic framework sketched in Table 8-1 is helpful for state leaders, themselves, in seeing, understanding, and valuing the role and place of balanced assessment systems, it may also serve as a potential resource in their efforts to craft a sense of coherence for others amidst what might otherwise present as fragmentation, incoherence, and turbulence.

For example, as summarized in Table 8-1, our analytic framework has policy-level priorities for advancing educational quality and equity as a 75-year through-line stretching back to the mid-20th century. Furthermore, the framework represents accumulating policy-level understandings of the support needed to advance educational quality and equity: more and better educational resources distributed more equitably among local public school districts; improving instructional practice and the school and district contexts in which it is situated; and developing and mobilizing local agency and capability, broadly and inclusively, to identify and address local educational ambitions, needs, and problems. Still further, the framework represents the comprehensive structural/procedural, technical, and moral press on local districts to advance educational quality and equity.

And, importantly, our analytic framework associates the central ideas of balanced assessment systems with these accumulating policy priorities and understandings, such that the central ideas of balanced assessment systems amplify the structural/procedural, technical, and moral press on local districts. This includes:

- integrated assessments as resources designed to serve different purposes among different actors at the classroom, local, district, and state levels, with a particular focus on learning theories that center the development of the whole child in community;
- the aim of coordinating the practice of organizing, managing, and improving instruction from the classroom level to the state level, with a particular focus on ambitious teaching; and
- the potential for comprehensive, coherent, and continuous assessment at all levels to provide the evidence needed not only to make incremental adaptations within existing systems but, also, to bring diverse perspectives to bear on interrogating, disrupting, and reforming existing systems.

With that, our analytic framework and Table 8-1 provide a blueprint for state leaders in crafting a coherent policy narrative that positions balanced assessment systems squarely within leading lines of reasoning and action aimed at advancing educational quality and equity. One matter is how such a narrative would be taken up by local districts (a matter that we discuss below, in our analysis of “seeing and crafting” at the local level). A more immediate matter is how such a coherent narrative would be taken up by their state-level colleagues responsible for other domains of education policy. After all, new ideas are always interpreted through and grafted onto existing understandings. From that follows the risk that existing understandings might shape new ideas more than new ideas reshape existing understandings. This is an essential finding in research on instructional innovations at the level of the individual teacher (for a seminal analysis, see Cohen, 1990). There is no reason to suspect anything different for the individual policy maker.

On the assumption that policy logics accumulate in the minds of individual policy makers much as in the national educational ecology in which they live and work, our analytic framework may be useful for state leaders in speculating about the inherited understandings through which the idea of balanced assessment systems will be interpreted and understood by their colleagues. For example:

- Many state-level colleagues may understand and work within a resource-forward logic, on the belief that more and better educational resources are sufficient to advance educational quality and equity. Such colleagues would be disposed to recognizing the *mechanisms* of balanced assessment systems.
- Some colleagues may go further to also understand and work within a practice-forward logic, on the assumption that standards, evidence, and accountability will be sufficient to evoke new behaviors in districts, schools, and classrooms that advance academic outcomes. Such colleagues would be positioned to recognize the *practical ambitions* of balanced assessment systems.
- Fewer colleagues are likely to understand and work within an empowerment-forward logic, with commitment to developing and mobilizing local agency and capability, broadly and inclusively, to identify and address local educational ambitions, needs, and problems that may go beyond (or reframe) academic outcomes. Such colleagues would be positioned to recognize the *theory of action* of balanced assessment systems.

Variation in inherited understandings, in turn, suggests a formidable challenge for state leaders in crafting coherence. As a precondition for understanding the idea of balanced assessment systems, state leaders will need to collaborate with their colleagues to explicate, reflect critically upon, and likely further develop the fundamental policy logics that structure their understandings and work. Absent efforts of this sort, the risk is that, however coherent the narrative, the idea of balanced assessment systems will be misunderstood among their state-level colleagues: for example, as new resources that warrant no more than structural/procedural compliance; as new expectations for improving professional practice, absent support for developing professional capabilities; and as new ambitions for advancing educational quality and equity, absent efforts to cultivate inclusion and agency among historically marginalized groups and communities.

### SEEING AND CRAFTING AT THE LOCAL LEVEL

In the United States, local public school districts shoulder primary responsibility for organizing, managing, and improving their essential educational work—instruction, teaching, and learning—in ways that mediate between the national education ecology and local ambitions for students’ educational opportunities, experiences, and outcomes. Our next step, thus, is to take up the matter of seeing systems and crafting coherence at the local level, with the aim of developing shared understandings of the place and role of balanced assessment systems among local efforts to advance educational quality and equity.

Unlike the national education ecology, local districts are more meaningfully examined as systems: again, collections of organizations intentionally designed and structured to work in interaction, with shared purpose and toward a common goal. But this is not to say that this “systemness” has local districts any less complicated than the national education ecology. Rather, as systems, they vary remarkably in form and gov-

ernance, including conventional geopolitical public school districts overseen by local school boards, charter school networks overseen by state-approved authorizing agencies, and “turnaround districts” overseen by state-appointed boards. Moreover, different districts have different combinations and configurations of elementary, middle, and/or high schools.

The conventional narrative is that policy-level fragmentation, incoherence, and turbulence has fostered fragmentation, incoherence, and turbulence within local districts. This narrative has central offices and schools attending to changing policy ambitions, priorities, and requirements that they see as most important and relevant to their work, each in their own locally sensible way.

But, again, as represented in Table 8-1, we have identified structure and order in ways that districts have co-evolved with the national education ecology: specifically, as *school* systems, *education* systems, and *learning* systems.<sup>15</sup> Each of these system types is characterized by different functional capabilities for organizing, managing, and improving instruction to advance educational quality and equity. Each is a response to an associated, policy-level focus on *resources*, *practice*, and *empowerment* and, with that, an associated press to maintain *structural/procedural*, *technical*, and *moral legitimacy*. As such, these system types form a taxonomy—a developmental progression—that frames the accumulation of functional capabilities in local public school districts in response to accumulating policy logics and legitimacies in the national education ecology.

Our central line of argument is that, by seeing and understanding this developmental progression from *school* systems to *education* systems to *learning* systems, state and local leaders will be better positioned to assess the current capabilities of local districts in advancing educational quality and equity; their capacity to engage the idea of balanced assessments systems less-or-more comprehensively; and ways in which fuller engagement with the idea of balanced assessments systems may require developing categorically distinct functional capabilities to organize, manage, and improve instruction.

### School Systems

As we define it, a *school system* is a local district distinguished by highly developed capabilities for organizing, managing, and improving *access to public schooling*, but comparatively weak capabilities for organizing, managing, and improving the *educational work* of teaching and learning once students are in schools. Our contention is that districts that have evolved only as school systems are most apt to recognize the mechanisms of balanced assessments systems and to engage through structural/procedural compliance.

The evolution of districts as school systems predates the increased policy focus on resource-forward innovation and improvement in the mid-20th century. By that point, local districts had been under a century-long societal and policy press to establish and institutionalize mass public schooling to increase access to instruction for more (and more diverse) students: for example, through the common schools movement, the introduction of compulsory attendance and truancy laws, mass immigration, and urbanization.

---

<sup>15</sup> For our conceptual development of school, education, and learning systems, see Datnow et al. (2022) and Peurach et al. (2019a, 2019b, 2022a).



From this century-long press for mass public schooling emerged the organizational template of what would become widely recognized, valued, and understood to be a local public school district (e.g., Callahan, 1964; Cuban, 1988; Tyack, 1974). Key characteristics included:

- the emergence of a central office to administer schools, staffed by professional administrators and accountable to a democratically elected local school board;
- structural differentiation between levels of schools (elementary, middle, and high schools) and within schools (grade levels, academic content areas, and academic tracks—college preparatory, general education, and vocational); and
- a conventional distribution of labor, with local central office and school leaders responsible for managing both political relationships and administrative requirements, and with teachers responsible for managing the educational work of classroom instruction.

From this century-long press for mass public schooling also emerged a conventional approach to organizing, managing, and improving instruction: one that we describe as *sorting*, *resourcing*, and *delegating* (Peurach et al., 2019b). Central office and school leaders provided access to instruction by sorting students into schools, grade levels, academic tracks, and classes. They resourced those instructional venues with teachers, curriculum frameworks, textbooks, and other instructional materials and guidance, supported primarily by local tax revenues. And they delegated to teachers primary responsibility for organizing and managing the day-to-day work of classroom instruction for the students assigned to them using the resources afforded them.

Indeed, the century-long emergence and institutionalization of a historically novel organizational form—local public school districts—across a rapidly growing, rapidly changing, and remarkably diverse country yielded an odd result: homogeneity. With much to predict variation in structure and practice, scholars instead described the emergence and institutionalization of “the grammar of schooling” and the “one best system,” with a “real school” being one that heeded this grammar (e.g., Metz, 1989; Tyack, 1974; Tyack & Cuban, 1997; Tyack & Tobin, 1994).

This comparatively strong focus on the structure of local public school districts was *cultural*, in that it came to be recognized, valued, and understood in ways just described; *functional*, in that it supported access to instruction for more (and more diverse) students; and *pragmatic* in multiple ways, including bringing public schooling into alignment with continuing education, career, and vocational opportunities for students as they exited. The comparatively weak focus on the educational work of classroom instruction owed much to local dynamics that paralleled those in the broader national education ecology: the lack of social agreement on the means and ends of instruction, and the political challenges and risks for local central office and school leaders in attempting to forge that agreement.

Beginning in the middle of the 20th century, policy-level priorities and initiatives in the national education ecology as described above (under “See and Crafting at the Policy Level”) did not disrupt these local-level dynamics: that is, the onset of increasing federal engagement in public education; the press for expanding equitable access to public schooling; the onset of resource-forward innovation and improvement aimed

at advancing educational quality and equity; and the press for structural/procedural legitimacy. To the contrary, these local dynamics were the very frame through which local public school districts apprehended and responded to those policy-level priorities and initiatives. For example:

- Local public school districts created new categories of schools (e.g., magnet schools) and new instructional venues (e.g., special education, Title I, second language, vision and hearing impaired, and gifted and talented) into which they sorted students whose education was newly prioritized and supported by federal funding.
- Local public school districts used new sources of federal and other discretionary funding to leverage the educational resource market to adopt new materials, programs, and services, with these new resources symbolizing engagement with the policy press to advance educational quality and equity.
- Absent agreement or accountability on matters of educational quality and equity, local public school districts maintained structural/procedural legitimacy by providing evidence of compliance with associated regulations and requirements, while continuing to delegate primary responsibility for day-to-day classroom instruction to teachers.

By the late 1970s and early 1980s, organizational researchers were describing local public school districts as “loosely coupled systems” rife with structures and resources, but with weak capabilities in central offices and schools to organize, manage, and improve the day-to-day work of teachers and students in classroom instruction (Meyer & Rowan, 1978). In that this arrangement was sufficient to maintain legitimacy and good standing, they also described it as a rational response to policy-level and local-level contexts that lacked agreement on the meaning and methods of advancing educational quality and equity.

### Education Systems

As we define it, an *education system* is a local district distinguished by capabilities among central offices and schools to collaborate with teachers to organize, manage, and improve the educational work of public schooling—instruction—with the aim of improving educational quality and reducing educational disparities. Districts heeding the press to evolve as educational systems are more apt to recognize the practical ambitions of balanced assessment systems and to engage in ways aimed at improving technical effectiveness (and not in ways that are simply structural and procedural).

In our past research, we associated the evolution of districts as educational systems with the onset of practice-forward innovation and improvement at the policy level and, with that, the press for technical legitimacy (Peurach et al., 2019b, 2022a). As sketched above, the onset of the practice-forward policy logic coincided with:

- the establishment of universal access to public schooling, thus institutionalizing the policy-level press on local educational enterprises to sustain as access-providing school systems;

- the accumulation of research providing increased transparency in the operations of local public school districts, including sorting, resourcing, and delegating as a root cause of low quality and inequitable educational opportunities, experiences, and outcomes for many students;
- the advent of an excellence and equity movement; the onset of new calls and ideas for ambitious teaching and learning; the introduction of policy initiatives advancing new ideas and priorities for systems thinking, standards and accountability, markets and choice, data and evidence, and autonomy and professionalism; and the consequent press on districts to improve instructional practice and its school/district contexts.

At the local level, one effect was to disrupt homogeneity in the structure and governance of public school districts, with the introduction and proliferation of charter school networks, within-district school choice, mayoral control, and state turnaround and takeover districts. As sketched above, another effect was increased dependence on the educational resource market for new (and new categories of) materials, programs, and services.

Yet another effect was the emergence of new patterns in the organization, management, and improvement of instruction. Beginning in the 2000s, new research examining the organization and operations of local public school districts detailed the emergence of new domains of activity distributed among central offices and schools that, when managed coherently and with coordination, supported teachers in improving instructional practice in ways that advanced educational quality and reduced educational disparities (e.g., Bryk et al., 2010; Cobb et al., 2018; Forman et al., 2017; Johnson et al., 2014). In a comprehensive review of this research, we summarized this new activity as five core domains of work enacted by central office and school leaders, often in collaboration with teachers (Peurach et al., 2019b). These domains include:

- *Managing environmental relationships* to discern, bridge, buffer, and reconcile the many influences bearing on how districts understand and pursue quality and equity in classroom instruction.
- *Building educational infrastructure* that coordinates visions for instructional practice, formal organizational resources (instructional models, curricula, routines, and assessments), and social organizational resources (norms, values, and relationships among teachers, leaders, and students).
- *Supporting and integrating the use of educational infrastructure in practice* by developing teachers' professional knowledge and capabilities through such means as workshops, practice-based coaching and mentoring, and collegial learning.
- *Monitoring and managing performance* both for continuous improvement (e.g., via iterative implementation, evaluation, and refinement of infrastructure and supports) and for accountability (e.g., via the use of evidence and standards to assess quality and equity in student outcomes).
- *Developing and distributing instructional leadership* beyond established administrative positions to new leadership roles and structures responsible for performing, coordinating, and managing the preceding domains of work.

We found nothing in our review (nor in our subsequent research) to suggest a tectonic shift in these directions within local districts, nor homogeneity in the ways that districts are taking up these core domains of work. While some districts are advancing comprehensive strategic plans, others are developing these capabilities more incrementally and organically. Moreover, efforts to develop these capabilities appear to be concentrated in academic content areas that are the primary focus of state standards, assessment, and accountability (English language arts, mathematics, and science), though with different urgency in different schools and levels of schooling (elementary, middle, and high schools) depending on district priorities. And districts appear to be developing, distributing, and coordinating these capabilities differently among central offices and schools in accord with different theories of action, as categorically distinct types of educational systems (Peurach et al., 2019b, 2019c).

### Learning Systems

As we define it, a *learning system* is a local district distinguished by capabilities to engage diverse stakeholders (professional, family, and community) in collaborating to develop the shared understandings, knowledge, and values needed to identify and address local educational ambitions, needs, and problems. Districts heeding the press to evolve as learning systems are more apt to recognize balanced assessment systems as supporting both incremental improvement and transformative change, and to engage from perspectives that are both technical and moral (and, again, not simply structural and procedural).

In our past research, we located the onset of this evolution toward learning systems at the intersection of practice-forward and empowerment-forward innovation and improvement, and in the interdependence between the press for technical and moral legitimacy (Peurach et al., 2019b, 2022a). As sketched above, the emergence of empowerment-forward innovation and improvement was motivated, in part, by research suggesting both (a) weaknesses in local capabilities to organize, manage, and improve instruction in response to new accountability demands and (b) potential to organize local districts in new ways to support continuous learning and improvement. Subsequently, it has been driven by a policy-level improvement movement aimed at developing capabilities and agency among diverse local stakeholders to use formal methods of collaborative, continuous improvement to understand and address local educational problems, needs, and opportunities, as well as by academic, intellectual, and social movements pressing for new approaches to equity and justice in the goals and work of educational improvement.

Local-level engagement is nascent: some self-initiated and self-guided, and more through grant-funded initiatives in association with external organizations (with leaders including the Strategic Education Research Partnership, the National Center on Scaling Up Effective Schools, the National Network of Education Research-Practice Partnerships, and the LearnDBIR initiative in the Research+Practice Collaboratory). Evidence of such engagement lies in the annual Carnegie Summit for Improvement in Education, which, since 2014, has supported thousands of participants (most from local districts) in developing capabilities for collaborative, continuous improvement. This includes local educational enterprises that have been identified as exemplars for

producing evidence associating the rigorous application of improvement methods with evidence of improving quality (and reducing disparities) in student outcomes (Bryk, 2020).

In our own research, we initially framed the evolution of districts as learning systems in terms of their emergence and development as “scientific-professional learning communities” that use rigorous methods of continuous improvement to produce, use, and refine the practical knowledge needed to advance educational quality and equity in locally responsive ways (Russell et al., 2017, 2019). Our initial efforts focused on a particular approach for enacting and organizing collaborative, continuous improvement: improvement science in networked improvement communities. We have since elaborated our framework as a general resource for examining the structures, norms, and capabilities essential for districts to productively function as learning systems. These structures, norms, and capabilities include:

- networked organizational structures and roles that connect and engage diverse teams (a) within and between local sites and (b) with partner organizations;
- a culture in which team members share a collective, evidence-based orientation to advance more equitable educational outcomes;
- capabilities to use formal methods of collaborative, continuous improvement to iteratively analyze systemic causes of educational weaknesses and disparities, to design and test interdependent interventions, and to evaluate effects on student outcomes;
- formal structures for collecting and exchanging data and for accumulating and managing practical knowledge;
- means of aligning and coordinating with other school and district initiatives; and
- leadership capabilities distributed among sites and partner organizations to structure, manage, and continuously improve the preceding structures, norms, and capabilities.

The preceding developments have played out in interaction with other policy-level dynamics that sit at the intersection of practice-forward and empowerment-forward innovation and improvement, including both (a) the rise of critical perspectives on practice-forward innovation and improvement and (b) the reinvigoration of a racial and social justice movement in education calling for participation of marginalized groups and communities in identifying and rectifying systemic inequities.

With the twin press to maintain both technical and moral legitimacy, some of these policy dynamics have led to the evolution of districts as education systems, including:

- the development of educational infrastructure that includes visions for holistic student development, culturally responsive pedagogies and curriculum materials, and norms of inclusion and mutual respect;
- the incorporation of equity audits into performance management; and
- the development of new leadership roles charged with advancing diversity, equity, inclusion, and anti-racism.



Other of these policy dynamics have led to the evolution of districts beyond scientific-professional learning communities to more democratic, inclusive learning systems. This evolution has been driven, in part, by complementary efforts at the policy level and the local level to bring the improvement movement into closer engagement with the racial and social justice movement. Examples include efforts by researchers to integrate commitments to diversity, equity, and inclusion into the principles and methods of continuous improvement; efforts to engage historically marginalized communities in using formal methods of continuous improvement to co-design educational interventions; the proliferation of research-practice partnerships that aim to integrate improvement and equity; and a landmark philanthropic effort—the Bill and Melinda Gates Foundation’s Networks for School Improvement Initiative—that has sought to integrate equity commitments into network-based improvement efforts to advance secondary school experiences and outcomes for students of poverty and color.<sup>16</sup>

### Crafting Coherence

Thus, there is much in our local-level analysis that plays directly to the conventional narrative, with differences in resources, structures, capabilities, and commitments between and within public school districts as artifacts of policy-level fragmentation, incoherence, and turbulence. There is much to suggest that there is more structure and order to the ways that districts organize, manage, and improve instruction than meets the eye. And there is much to suggest that this structure and order will serve as a frame through which districts apprehend and making meaning of the idea of balanced assessment systems.

Even if state leaders craft a coherent narrative that “gives sense” to the role and place of balanced assessment systems among other policy-level initiatives seeking to advance educational quality and equity, much depends on how local districts make sense of that message among many others. Organizational scholars would describe this as local districts “enacting” their policy contexts: looking beyond their boundaries; noticing some things; failing to notice others; intentionally ignoring still others; and making interpretations and judgments about what that they notice means and how they should act in relation (Weick, 1979).

How districts, as organizations, make sense of and respond to their policy contexts is, in part, a function of their current capabilities. Organizational scholars describe this in terms of districts’ “absorptive capacity”: their ability to recognize the meaning and value of new information, incorporate it into current understandings and ways of working, and act on it (Cohen & Levinthal, 1990; Farrell & Coburn, 2017). More commonly, the relationship between absorptive capacity and current capabilities is the gist of what is known as the Matthew effect: Those with advantage readily accumulate more advantage at an increasingly rapid rate, while those who lack advantage struggle to accumulate more advantage and become increasingly disadvantaged by comparison. How districts make sense of and respond to their policy contexts is also linked to frameworks describing the likelihood that organizations will adopt innovations: early

---

<sup>16</sup> For efforts that bring the improvement movement into closer dialogue with the racial and social justice movement, see Bush-Mecenas (2022); Ghiso et al. (2022); Hinnant-Crawford (2020); Ishimaru et al. (2019); and Peurach et al. (2022a). For specific analyses associating *learning systems* with commitments to empowerment, see Yurkofsky et al. (2020).

adopters having dispositions to search for novel tools, resources, and ideas, along with the capabilities to assimilate and leverage them; majority adopters (early and late), less so; and laggards, even less so (Rogers, 2010).

By this line of reasoning, how district leaders apprehend and make meaning of the idea of balanced assessment systems will be a function, in part, of their current capabilities for organizing, managing, and improving instruction: that is, where they lie along the developmental progression from *school system* to *education system* to *learning system*.

- Districts that are evolving as learning systems are positioned to perceive and understand the idea of balanced assessment systems at its fullest: its mechanisms, practical ambitions, and theory of action. Such districts are also positioned to engage in ways that establish their structural/procedural, technical, and moral legitimacy.
- Districts that are evolving (or that have evolved) as education systems are positioned to partially perceive and understand the mechanisms and practical ambitions of balanced assessment systems, and possibly its notions of incremental improvement within existing systems. Such districts are also positioned to engage in ways that establish their structural/procedural legitimacy and technical legitimacy.
- Districts that have only evolved as school systems are positioned to perceive and understand only the mechanisms of balanced assessment systems. Such districts are also positioned to engage in ways that establish their structural/procedural legitimacy.

The distribution of districts among these categories is not clear. However, in that the policy dynamics motivating their evolution vary from institutionalized to emergent (and from centuries old to years old), one conjecture would be that there would be few that are evolving as learning systems, some that are evolving as education systems, and some that have not evolved beyond school systems.

For state leaders charged with crafting a coherent introduction to the idea of balanced assessment systems, our analytic framework as summarized in Table 8-1 suggests advantage in complementing the type of policy-level narrative sketched above (in “Seeing and Crafting at the Policy Level”) with differentiated messaging that anticipates predictable variation in local-level sensemaking. Our analytic framework also suggests advantage in going further, by couching any messaging about balanced assessment systems in visions for the progressive development of capabilities in districts that would enable them to respond more fully to policy-level ambitions for advancing educational quality and equity.

For district leaders charged both with making sense of the idea of balanced assessment systems and with crafting a coherent, local-level interpretation of their role and place in the district, this local-level framework of *school*, *education*, and *learning* systems has at least four potential uses.

- It can serve as a framework through which to interpret the idea of balanced assessment systems more richly, in ways that transcend the level of extant capabilities in the district.

- It can serve as a diagnostic tool for self-analyzing the district’s current level of capabilities to organize, manage, and improve instruction to advance educational quality and equity—and, thus, its readiness to engage the idea of balanced assessment systems (as well as the level at which it is ready to engage).
- It can serve as a framework for envisioning the further development of the districts’ capabilities to advance educational quality and equity and, with that, its potential for higher levels of engagement with the idea of balanced assessment systems.
- It can serve as a resource for developing, implementing, and institutionalizing other complex educational innovations, ranging from multi-tiered systems of supports in elementary schools to graduation “on track indicators” in high schools.

## **LEARNING WHILE LEADING**

Our focus on introducing the very idea of balanced assessment systems has taken our analysis some distance away from conventional understandings about developing, implementing, and institutionalizing educational innovations. As such, we continue with a fuller discussion of what the work of “seeing and crafting” implies for state and local leaders. Specifically, we explore the need for state and local leaders to engage in two challenging tasks simultaneously: *learning* to see systems and to craft coherence while, at the same time, actually *leading* the work of introducing balanced assessment systems in their respective contexts.

### **The Learning Imperative**

Our analysis in the two preceding sections has state and local leaders orchestrating individual-level and organizational-level social learning processes that bring a complex, systemic innovation (balanced assessment systems) into engagement with institutionalized-but-evolving understandings of educational policy, organization, and practice. These learning processes have state and district leaders needing to develop shared schema among colleagues and in organizations to apprehend (and to give meaning to) policy and reform activity as it has accumulated (and continues to accumulate) at the state and local levels. That, in turn, requires explicating existing schema, reconciling differences between the new and the old, and calibrating expectations appropriately and developmentally.

Our analysis suggests that engaging in these learning processes is essential for fully understanding the meaning and place of balanced assessment systems in advancing educational quality and equity. Absent such work, the greatest risk, in our view, is that the idea of balanced assessment systems will be apprehended and taken up only within the deeply institutionalized status quo: as resources that require structural and procedural compliance, thus engaged symbolically and ritualistically—with little bearing on practice, and with little critical examination of the systems in which practice is situated (Peurach et al., 2018, 2019b; Yurkofsky, 2020).

Our prior research on building and rebuilding systems predicts that the work of seeing systems and crafting coherence will be exceedingly complex, and rife with dilem-

mas and paradoxes that complicate charting clear paths forward (Cohen et al., 2014, 2018; Peurach et al., 2019a; Seeber et al., under review; Yurkofsky & Peurach, 2023). Chief among these dilemmas and paradoxes is what organizational scholars describe as “the paradox of embedded agency”: State and local leaders are products, inhabitants, and stewards of the very systems that they aim to improve, thus constrained cognitively, practically, and politically in imagining and pursuing alternatives (Garud et al., 2007).

Managing these dilemmas and overcoming the paradox of embedded agency will challenge state and district leaders to develop opportunities for *their own collegial learning* as they work through the inevitable false starts, variable uptake, and difficult-to-discern progress endemic to building shared meaning and understanding of such a complex idea as balanced assessments systems both among themselves and in complex policy-level and local-level contexts. Again, this type of sensemaking and sensegiving—conducted iteratively, in the context of practical work—is, fundamentally, a social learning process.

To explore the challenge of learning while leading, we continue by drawing on two approaches to constructing and supporting leaders’ social, practice-based learning. The first is the development of a “learning lattice” for educational leaders that coordinates horizontal and vertical learning opportunities between and within states and districts. The second is the development of “learning systems” as discussed above: that is, of collaborative, continuous learning and improvement in network contexts.

### The Learning Lattice

Despite the learning imperative as discussed above, the history of lackluster results from educational reforms in the United States converges on a common theme: Reforms seeking to improve teaching and learning often fail to produce the changes in educator understanding and practice needed to achieve their aims. This is, in part, a result of educators at all levels of the system not having sufficient opportunities to learn the new practices required of the reform.<sup>17</sup>

Again, if balanced assessment systems are to break from this pattern, state and local leaders charged with developing, implementing, and institutionalizing them will need opportunities to learn about balanced assessment systems while, at the same time, orchestrating learning opportunities for others at the state, district, and school levels. This is no small challenge, given the complexity of the idea of balanced assessment systems: a reform that requires a high degree of coordination and interdependence in the understandings and practices of educators at the state, district, school, and classroom levels.

Taking inspiration from efforts to build leadership capacity for school-level improvement in England, Jonathan Supovitz builds a vision for a “learning lattice” that integrates formal and social learning opportunities for school leaders (Supovitz, 2014). Within this design, a central leadership development program creates and supports formal learning opportunities, while lateral social networks among leaders create ongoing opportunities for idea exchange, mentorship, and collaborative problem solving.

---

<sup>17</sup> For leading research on the learning demands of instructionally focused policy, see Cohen and Barnes (1993); Cohen and Hill (2001); Hubbard et al. (2006); Spillane (2000); and Stein and Coburn (2008).

The learning lattice is enacted within a broader framework for school leadership that provides clear guidance for leaders at multiple levels of organization, including school principals, assistant principals, and department chairs or lead teachers.

We find this notion of a learning lattice relevant as we think about the learning demands of balanced assessment systems, though we extend the design to include leadership learning opportunities at the state and district levels. We imagine a role for intermediary organizations or university-based partners with specific expertise to create formal learning opportunities about balanced assessment systems that are made broadly available through massive open online courses or other online learning formats. Complementing formal learning opportunities, we see the need to focus on the development of more informal or lateral connections among educators to support them in learning while leading.

### **A Lattice of Learning Opportunities for State Leaders**

We can imagine state leaders benefitting from opportunities to learn collaboratively with other leaders charged with developing, implementing, and institutionalizing balanced assessment systems, as well as cross-state learning opportunities for exchanging ideas with colleagues in other contexts. We also imagine such learning opportunities as a context in which state leaders could leverage the policy-level framework sketched in Table 8-1 to structure and inform their collaborative learning.

A research-practice partnership emerging from the work of the English Learner Collaborative of the Council for Chief State School Officers (CCSSO) presents a compelling model for organizing this kind of learning opportunity. State education agency (SEA) professionals are charged with implementing state and federal policy, dispersing resources, and providing guidance for schools related to the education of multilingual learners. Their learning needs emerged in the CCSSO's English Learner Collaborative and led to the establishment of a research-practice partnership to build capacity to promote equity for multilingual learners (Hopkins et al., 2022; Weddle et al., under review).

Coordinated through a research-practice partnership, this learning community includes 20 SEA leaders and their research partners from across the country. The partnership has been meeting every 3 weeks for 2.5 years in the form of small groups organized around relevant problems of practice, such as strategizing about the most productive use of multilingual learner funds and developing supports for multilingual learners with disabilities. Researchers share ideas about evidence-based practice, such as the need to foster shared responsibilities for multilingual learner success across the state, districts, and schools, which are then taken up in the collaborative work of small groups. Overall, the researchers report that SEA professionals are hungry for connections with peers, with the partnership filling a need for social learning that is not readily available in their state agency contexts.

We see great potential in the formation of similar research-practice partnerships as learning communities that support SEA professionals charged with catalyzing the movement toward balanced assessment systems. Several existing initiatives appear to present opportunities: for example, the CCSSO's Balanced Assessment System State Collaborative, the Technical Issues in Large-Scale Assessment (TILSA) Collaborative, and the Chief Academic Office Collaborative. However, these are typically three times



per year convenings, and not the frequent meetings of the Early Learning Collaborative's learning community as coordinated within a research-practice partnership. Refashioning these existing collaboratives as rigorous research-practice partnerships that could support the learning of state leaders would require commitment to more regular engagement in structured collaborative learning.

### **A Lattice of Learning Opportunities for Local Leaders**

As state leaders advance their own learning, we imagine them also coordinating networks of districts working collaboratively to advance balanced assessment systems. For inspiration, we point to the case of the Tennessee Mathematics Coaching Project (Russell et al., 2017). Recognizing a need to expand learning opportunities for teachers related to the implementation of ambitious mathematics standards, state leaders in Tennessee partnered with researchers and professional development providers to create a network of instructional leaders (i.e., coaches) in districts throughout the state. The network created structured learning opportunities that coaches took back to their local sites of practice and shared with local colleagues. Additionally, coaches systematically tested ways to integrate evidence-based coaching practices into their local systems and shared what they were learning with the network (Russell et al., 2020). As the project converged on a model for mathematics instructional coaching that could be implemented at scale, the state organized structured learning opportunities for district leaders through its regional units.

Other states could similarly convene networks of district leaders to develop a shared vision for balanced assessment practice and to grapple with the problems of practice inherent with a move toward this vision. Similarly, we can envision districts coordinating networks of schools working collaboratively to develop strategies for implementing balanced assessment systems.

Regarding district-coordinated networks, a helpful model is the work of the Baltimore Secondary Literacy Improvement Community, which is organized by the Baltimore City Schools.<sup>18</sup> The district convenes a network of "teaching fellows" who collaborate across schools to identify new ways to rapidly improve students' reading skills. Patterned after the networked improvement community model, district leaders create structured opportunities for teachers to test ways to develop secondary students' reading competencies. Teachers representing teams in schools throughout the district meet as working groups that tackle specific components of the literacy problem (e.g., fluency, vocabulary). In network convenings and through strategic knowledge management practices, teachers share what they are learning so that promising practices can spread across schools.

We imagine networks of this sorts as contexts in which local leaders, themselves, could learn about (and from) the work of crafting a coherent vision for the introduction of balanced assessment systems, observing teachers as they enact and operationalize that vision, and working iteratively with teachers to refine both the vision and its enactment.

---

<sup>18</sup> See <https://www.baltimorecityschools.org/progress-BSLIC-fellows>.

## KEY TAKEAWAYS

Our view is that success developing, implementing, and institutionalizing balanced assessment will depend on orchestrating the learning of state and local leaders, beginning with developing shared understandings of the very idea of balanced assessment systems and of their place among other initiatives aiming to advance educational quality and equity. Such learning benefits from a comprehensive, historical perspective on the co-evolution of national policy contexts and local districts, as well as from insights into novel approaches to organizing collaborative, practice-based social learning. Our view holds for any comprehensive, practice-focused systems improvement initiative aiming to advance educational quality and equity through instructional improvement.

We are not alone in calling for increased attention to the learning demands of balanced assessment systems. Rather, this call runs thick through the other chapters in this volume, which emphasize the need for clear, coherent communications and messaging about the ideas of balanced assessment systems; the need to develop professional learning infrastructures and resources; the potential benefits of research-practice partnerships; and the need to establish an agenda and climate for learning in districts and schools.

The promising news is that a variety of models have emerged in the education field for organizing collective learning and improvement to drive the understanding and enactment of complex, distributed, instructionally focused reforms. The cases that we present in our discussion of “Learning While Leading” immediately above share two common features. First, implementation is viewed as a learning problem, and leaders at the state, district, and school levels engaging in collaborative problem solving. The novel emphasis is on building lateral connections among educational professionals who have not often had opportunities to work with role-alike leaders in other contexts. Second, each of these cases benefits from collaboration with external research partners.

Federal and philanthropic investment in the development of research partners to support such learning processes has thus far been weak as compared to other research investments (Peurach et al., 2018). While a scarce resource, such research partners have potential to provide analytical capacity to productively integrate evidence-based practices with practitioner knowledge in ways needed to understand and implement complex, distributed, instructionally focused innovation and improvement.

This, in our view, is the frontier of U.S. education reform: empowering diverse stakeholders to collaborate in advancing educational quality and equity through the development of inclusive, evidence-based, practice-focused learning systems. For proponents of balanced assessment systems, the most fundamental challenge is to collaborate with state and district leaders to create for themselves exactly the types of learning systems that they must strive to create for others.

## REFERENCES

- Allington, R. L., & Johnston, P. (1989). Coordination, collaboration, and consistency: The redesign of compensatory and special education interventions. In R. E. Slavin, N. L. Karweit, & N. A. Madden (Eds.), *Effective programs for students at risk* (pp. 320–354). Allyn and Bacon. <https://files.eric.ed.gov/fulltext/ED332290.pdf>.
- Au, W. (2010). *Unequal by design: High-stakes testing and the standardization of inequality*. Routledge, Taylor & Francis Group.

- Bryk, A. S. (2020). *Improvement in action: Advancing quality in America's schools*. Harvard Education Press.
- Bryk, A. S., Sebring, P. B., Kerbow, D., Rollow, S., & Easton, J. Q. (1999). *Charting Chicago school reform: Democratic localism as a lever for change*. Westview Press.
- Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing school for improvement: Lessons from Chicago*. The University of Chicago Press.
- Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard University Press.
- Burch, P. (2009). *Hidden markets: The new education privatization*. Routledge.
- Bush-Mecenas, S. (2022). "The business of teaching and learning": Institutionalizing equity in educational organizations through continuous improvement. *American Educational Research Journal*, 59(3), 461–499.
- Callahan, R. E. (1964). *Education and the cult of efficiency: A study of the social forces that have shaped the administration of the public schools*. The University of Chicago Press.
- Cobb, P., Jackson, K., Henrick, E., Smith, T. M., & the MIST Team. (2018). *Systems for instructional improvement: Creating coherence from the classroom to the district office*. Harvard Education Press.
- Coburn, C. E., Hill, H. C., & Spillane, J. P. (2016). Alignment and accountability in policy design and implementation: The Common Core State Standards and implementation research. *Educational Researcher*, 45(4), 243–251. <https://doi.org/10.3102/0013189X16651080>.
- Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis*, 12(3), 311–329. <https://doi.org/10.3102/01623737012003311>.
- Cohen, D. K., & Ball, D. L. (1999). *Instruction, capacity, and improvement*. CPRE Research Report Series RR-43. [https://repository.upenn.edu/cpre\\_researchreports/8](https://repository.upenn.edu/cpre_researchreports/8).
- Cohen, D. K. & Barnes, C. A. (1993). Pedagogy and policy. In D. K. Cohen, M. W. McLaughlin, & J. E. Talbert (Eds.), *Teaching for understanding: Challenges for policy and practice* (pp. 207–239). Jossey-Bass.
- Cohen, D. K. & Hill, H. C. (2001). *Challenges for policy and practice*. Jossey-Bass Inc.
- Cohen, D. K. & Mehta, J. D. (2017). Why reform sometimes succeeds: Understanding the conditions that produce reforms that last. *American Educational Research Journal*, 54(4), 644–690. <https://doi.org/10.3102/0002831217700078>.
- Cohen, D. K., & Spillane, J. P. (1992). Chapter 1: Policy and practice: The relations between governance and instruction. *Review of Research in Education*, 18(1), 3–49. <https://doi.org/10.3102/0091732X018001003>.
- Cohen, D. K., McLaughlin, M. W., & Talbert, J. E. (1993). *Teaching for understanding: Challenges for policy and practice*. Jossey-Bass Publishers.
- Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, 25(2), 119–142. <https://doi.org/10.3102/01623737025002119>.
- Cohen, D. K., Peurach, D. J., Glazer, J. L., Gates, K. E., & Goldin, S. (2014). *Improvement by design: The promise of better schools*. The University of Chicago Press.
- Cohen, D. K., Spillane, J. P., & Peurach, D. J. (2018). The dilemmas of educational reform. *Educational Researcher*, 47(3), 204–212. <https://doi.org/10.3102/0013189X17743488>.
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1), 128–152. <https://doi.org/10.2307/2393553>.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. U.S. Government Printing Office. <https://files.eric.ed.gov/fulltext/ED012275.pdf>.
- Cuban, L. (1988). *The managerial imperative and the practice of leadership in schools*. State University of New York Press.
- Datnow, A., Park, V., Peurach, D. J., & Spillane, J. P. (2022). *Transforming education for holistic student development: Learning from education system (re)building around the world*. The Brookings Institution.
- Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37(1), 15–24.
- Elmore, R. F. (2002). *Bridging the gap between standards and achievement: The imperative for professional development in education*. Albert Shanker Institute. <https://www.shankerinstitute.org/resource/bridging-gap-between-standards-and-achievement>.
- Farrell, C. C., & Coburn, C. E. (2017). Absorptive capacity: A conceptual framework for understanding district central office learning. *Journal of Educational Change*, 18, 135–159. <https://doi.org/10.1007/s10833-016-9291-7>.

- Forman, M. L., Stosich, E. L., & Bocala, C. (2017). *The internal coherence framework: Creating the conditions for continuous improvement in schools*. Harvard Education Press.
- Fuhrman, S. H. (1993). *Designing coherent education policy: Improving the System*. Jossey-Bass Publishers. <https://files.eric.ed.gov/fulltext/ED359626.pdf>.
- Fuller, B., & Kim, H. (2022). *Systems thinking to transform schools: Identifying levers that lift educational quality*. The Brookings Institution. [https://www.brookings.edu/wp-content/uploads/2022/08/Brookings\\_Brief\\_Systems-thinking-to-transform-schools\\_v13.pdf](https://www.brookings.edu/wp-content/uploads/2022/08/Brookings_Brief_Systems-thinking-to-transform-schools_v13.pdf).
- Garud, R., Hardy, C., & Maguire, S. (2007). Institutional entrepreneurship as embedded agency: An introduction to the special issue. *Organization Studies*, 28(7), 957–969. <https://doi.org/10.1177/0170840607078958>.
- Ghiso, M. P., Campano, G., Thakurta, A., & Ponce, O. V. (2022). Community-based research with immigrant families: Sustaining an intellectual commons of care, resistance, and solidarity in an urban intensive context. *Urban Education*. <https://doi.org/10.1177/00420859221082676>.
- Glazer, J. L., & Peurach, D. J. (2015). Occupational control in education: The logic and leverage of epistemic communities. *Harvard Educational Review*, 85(2), 172–202. <https://doi.org/10.17763/0017-8055.85.2.172>.
- Hess, F. M. (1998). *Spinning wheels: The politics of urban school reform*. Brookings Institution Press.
- Hodge, E. M., & Stosich, E. L. (2022). Accountability, alignment, and coherence: How educators made sense of complex policy environments in the Common Core era. *Educational Evaluation and Policy Analysis*, 44(4), 543–566. <https://doi.org/10.3102/01623737221079650>.
- Hodge, E. M., Salloum, S. J., & Benko, S. L. (2019). The changing ecology of the curriculum marketplace in the era of the Common Core State Standards. *Journal of Educational Change*, 20, 425–446. <https://doi.org/10.1007/s10833-019-09347-1>.
- Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher*, 33(8), 16–30. <https://doi.org/10.3102/0013189X033008016>.
- Hopkins, M., Weddle, H., Castillo, M., Costa, J., Edwards, K., Elliot, S., Gautsch, L., Lowenhaupt, R., & Salas, V. (2022). Upholding multilingual learners' civil rights under ESSA: State education agency leaders and the contextual factors shaping their work. *American Journal of Education*, 128(4). <https://doi.org/10.1086/720362>.
- Hubbard, L. A., Stein, M. K., & Mehan, H. (2006). *Reform as learning: School reform, organizational culture, and community politics in San Diego*. Routledge.
- Ishimaru, A. M., Bang, M., Valladares, M. R., Nolan, C. M., Tavares, H., Rajendran, A., & Chang, K. (2019). *Recasting families and communities as co-designers of education in tumultuous times*. National Education Policy Center. <http://nepc.colorado.edu/publication/family-leadership>.
- Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America*. Harper Colophon Books.
- Johnson, S. M., Marietta, G., Higgins, M. C., Mapp, K. L., & Grossman, A. S. (2014). *Achieving coherence in district improvement: Managing the relationship between the central office and schools*. Harvard Education Press.
- Medin, D., Lee, C. D., & Bang, M. (2014). Point of view affects how science is done: Gender and culture influence research on a fundamental level. *Scientific American*, 311(4), 44–45. <https://doi.org/10.1038/scientificamerican1014-44>.
- Metz, M. H. (1989). Real school: A universal drama amid disparate experience. In D. E. Mitchell & M. E. Goertz (Eds.), *Education politics for the new century: The twentieth anniversary yearbook of the Politics of Education Association* (pp. 75–91). The Falmer Press. <https://files.eric.ed.gov/fulltext/ED319140.pdf>.
- Meyer, J. W., & Rowan, B. (1978). The structure of educational organizations. In M. W. Meyer (Ed.), *Schools and society: A sociological approach to education* (pp. 217–225). Jossey-Bass.
- National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. The National Academies Press. <https://doi.org/10.17226/24783>.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. U.S. Government Printing Office. [http://edreform.com/wp-content/uploads/2013/02/A\\_Nation\\_At\\_Risk\\_1983.pdf](http://edreform.com/wp-content/uploads/2013/02/A_Nation_At_Risk_1983.pdf).
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academy Press. <https://doi.org/10.17226/9853>.
- Oakes, J. (1985). *Keeping track: How schools structure inequality*. Yale University Press.



- Payne, C. M. (2008). *So much reform, so little change: The persistence of failure in urban public schools*. Harvard Education Press.
- Peurach, D. J. (2011). *Seeing complexity in public education: Problems, possibilities, and success for all*. Oxford University Press.
- Peurach, D. J., & Glazer, J. L. (2012). Reconsidering replication: New perspectives on large-scale school improvement. *Journal of Educational Change*, 13, 155–190. <https://doi.org/10.1007/s10833-011-9177-7>.
- Peurach, D. J., Glazer, J. L., & Winchell Lenhoff, S. (2016). The developmental evaluation of school improvement networks. *Educational Policy*, 30(4), 606–648. <https://doi.org/10.1177/0895904814557592>.
- Peurach, D. J., Penuel, W. R., & Russell, J. L. (2018). Beyond ritualized rationality: Organizational dynamics of instructionally-focused continuous improvement. In M. Connolly, D. H. Eddy-Spicer, C. James, & S. D. Kruse (Eds.), *The SAGE handbook of school organization* (pp. 465–488). Sage.
- Peurach, D. J., Cohen, D. K., & Spillane, J. P. (2019a). Governments, markets, and instruction: Considerations for cross-national research. *Journal of Educational Administration*, 57(4), 393–410. <https://doi.org/10.1108/JEA-09-2018-0172>.
- Peurach, D. J., Cohen, D. K., Yurkofsky, M. M., & Spillane, J. P. (2019b). From mass schooling to educational systems: Changing patterns in the organization and management of instruction. *Review of Research in Education*, 43(1), 32–67. <https://doi.org/10.3102/0091732X18821131>.
- Peurach, D. J., Yurkofsky, M. M., & Sutherland, D. H. (2019c). Organizing and managing for excellence and equity: The work and dilemmas of instructionally focused education systems. *Educational Policy*, 33(6), 812–845. <https://doi.org/10.1177/0895904819867267>.
- Peurach, D. J., Foster, A. T., Lyle, A. M., & Seeber, E. R. (2022a). Democratizing educational innovation and improvement: The policy contexts of improvement research in education. In D. J. Peurach, J. L. Russell, L. Cohen-Vogel, & W. Penuel (Eds.), *The foundational handbook on improvement research in education* (pp. 211–240). Rowman & Littlefield.
- Peurach, D. J., Russell, J. L., Cohen-Vogel, L., & Penuel, W. (2022b). *The foundational handbook on improvement research in education*. Rowman & Littlefield.
- Polikoff, M. S. (2012a). The association of state policy attributes with teachers' instructional alignment. *Educational Evaluation and Policy Analysis*, 34(3), 278–294. <https://doi.org/10.3102/0162373711431302>.
- Polikoff, M. S. (2012b). Instructional alignment under No Child Left Behind. *American Journal of Education*, 118(3), 341–368. <https://doi.org/10.1086/664773>.
- Polikoff, M. S. (2015). How well aligned are textbooks to the Common Core Standards in mathematics? *American Educational Research Journal*, 52(6), 1185–1211. <https://doi.org/10.3102/0002831215584435>.
- Powell, A. G., Farrar, E., & Cohen, D. K. (1985). *The shopping mall high school: Winners and losers in the educational marketplace*. Houghton Mifflin.
- Purkey, S. C., & Smith, M. S. (1983). Effective schools: a review. *The Elementary School Journal*, 83(4). <https://doi.org/10.1086/461325>.
- Purkey, S. C., & Smith, M. S. (1985). School reform: The district policy implications of the effective schools literature. *The Elementary School Journal*, 85(3), 352–389.
- Reckhow, S. (2012). *Follow the money: How foundation dollars change public school politics*. Oxford University Press.
- Rogers, E. M. (2010). *Diffusion of innovations* (4th ed.). The Free Press.
- Rowan, B. (2002). The ecology of school improvement: Notes on the school improvement industry in the United States. *Journal of Educational Change*, 3, 283–314. <https://doi.org/10.1023/A:1021277712833>.
- Russell, J. L., & Bray, L. E. (2013). Crafting coherence from complex policy messages: Educators' perceptions of special education and standards-based accountability policies. *Education Policy Analysis Archives*, 21(12). <https://doi.org/10.14507/epaa.v21n12.2013>.
- Russell, J. L., Meredith, J., Childs, J., Stein, M. K., & Prine, D. W. (2015). Designing inter-organizational networks to implement education reform: An analysis of state Race to the Top applications. *Educational Evaluation and Policy Analysis*, 37(1), 92–112. <https://doi.org/10.3102/0162373714527341>.
- Russell, J. L., Stein, M. K., Correnti, R., Bill, V., Booker, L., & Schwartz, N. (2017a). *Tennessee scales up improved math instruction through coaching*. [https://nasbe.nyc3.digitaloceanspaces.com/2019/10/TN-Math-Coaching-Pages-23-27\\_NASBE-Standard-May17\\_Final.pdf](https://nasbe.nyc3.digitaloceanspaces.com/2019/10/TN-Math-Coaching-Pages-23-27_NASBE-Standard-May17_Final.pdf).



- Russell, J. L., Bryk, A. S., Dolle, J., Gomez, L. M., Lemahieu, P. G., & Grunow, A. (2017b). A framework for the initiation of networked improvement communities. *Teachers College Record*, 119(5), 1–36. <https://doi.org/10.1177/016146811711900501>.
- Russell, J. L., Bryk, A. S., Peurach, D. J., Sherer, D., Khachatryan, E., LeMahieu, P. G., Sherer, J. Z., & Hannan, M. (2019, April). The social structure of networked improvement communities: Cultivating the emergence of a scientific-professional learning community. <https://files.eric.ed.gov/fulltext/ED616986.pdf>.
- Russell, J. L., Correnti, R., Stein, M. K., Bill, V., Hannan, M., Schwartz, N., Booker, L. N., Pratt, N. R., & Matthis, C. (2020). Learning from adaptation to support instructional improvement at scale: Understanding coach adaptation in the TN Mathematics Coaching Project. *American Educational Research Journal*, 57(1), 148–187. <https://doi.org/10.3102/0002831219854050>.
- Seeber, E. R., Spillane, J. P., Yin, X., Haverly, C., & Quan, W. (under review). *Leading systemwide improvement in elementary science education: Managing dilemmas of education system-building*. University of Michigan.
- Smith, M. S., & O'Day, J. (1990). Systemic school reform. *Journal of Education Policy*, 5(5), 223–267. <https://doi.org/10.1080/02680939008549074>.
- Spillane, J. P. (2000). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and Instruction*, 18(2), 141–179.
- Spillane, J. P. & Sun, J. M. (2020). The press for technical rationality & dilemmas of professional practice: Managing education in a pluralistic institutional environment. In L. Moos, E. Nihlfors, & J. M. Paulsen (Eds.), *Re-centering the critical potential of Nordic school leadership research* (pp. 71–87). Springer.
- Spillane, J. P., & Thompson, C. L. (1997). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. *Educational Evaluation and Policy Analysis*, 19(2), 185–203. <https://doi.org/10.3102/01623737019002185>.
- Spillane, J. P., Peurach, D. J., & Cohen, D. K. (2019a). Comparatively studying educational system (re)building cross-nationally: Another agenda for cross-national educational research? *Educational Policy*, 33(6), 916–945. <https://doi.org/10.1177/0895904819867264>.
- Spillane, J. P., Seelig, J. L., Blaushild, N. L., Cohen, D. K., & Peurach, D. J. (2019b). Educational system building in a changing educational sector: Environment, organization, and the technical core. *Educational Policy*, 33(6), 846–881. <https://doi.org/10.1177/0895904819866269>.
- Spillane, J. P., Blaushild, N. L., Neumerski, C. M., Seelig, J. L., & Peurach, D. J. (2022). Striving for coherence, struggling with incoherence: A comparative study of six educational systems organizing for instruction. *Educational Evaluation and Policy Analysis*, 44(4), 567–592. <https://doi.org/10.3102/01623737221093382>.
- Stein, M. K., & Coburn, C. E. (2008). Architectures for learning: A comparative analysis of two urban school districts. *American Journal of Education*, 114(4). <https://doi.org/10.1086/589315>.
- Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *The Academy of Management Review*, 20(3), 571–610. <https://doi.org/10.2307/258788>.
- Supovitz, J. (2014). *Building a lattice for school leadership: The top-to-bottom rethinking of leadership development in England and what it might mean for American education*. CPRE Research Reports. <https://repository.upenn.edu/entities/publication/f389470c-9dd1-473c-9a71-3fa38bc32337>.
- Tyack, D. (1974). *The one best system: A history of American urban education*. Harvard University Press.
- Tyack, D. & Cuban, L. (1997). *Tinkering toward utopia: A century of public school reform*. Harvard University Press.
- Tyack, D., & Tobin, W. (1994). The “grammar” of schooling: Why has it been so hard to change? *American Educational Research Journal*, 31(3), 453–479. <https://doi.org/10.3102/00028312031003453>.
- Vinovskis, M. A. (1999). *The road to Charlottesville: The 1989 Education Summit*. National Education Goals Panel.
- Weddle, H. (under review). Engaging in research-practice partnerships with policy intermediaries: A case study of collaboration with state education agency leaders. University of Pittsburgh.
- Weick, K. E. (1979). *The social psychology of organizing*. Addison-Wesley Publishing Company.
- Weick, K. E. (1995). *Sensemaking in organizations*. Sage Publications.
- Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization Science*, 16(4), 409–421.

- Yurkofsky, M. M., & Peurach, D. J. (2023). The paradox of leading amidst uncertainty: Maintaining balance on an unstable beam. *Journal of Educational Administration*, 61(3), 185–204. <https://doi.org/10.1108/JEA-09-2022-0168>.
- Yurkofsky, M. M., Peterson, A. J., Mehta, J. D., Horwitz-Willis, R., & Frumin, K. M. (2020). Research on continuous improvement: Exploring the complexities of managing educational change. *Review of Research in Education*, 44(1), 403–433. <https://doi.org/10.3102/0091732X20907363>.

# 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 .....	274
THE ROLE OF ASSESSMENT POLICY IN SUPPORTING TEACHING AND LEARNING .....	274
Policy-Related Purposes of Assessment.....	275
Federal- and State-Level Policy Making .....	276
Other Policy Influences and Actors .....	280
How Policies Are Designed to Shape Instruction, Learning, and Assessment.....	284
Research on Educator Responses to Assessment Policy.....	285
An International Perspective on Assessment Policy .....	287
UNDERSTANDING AND ADDRESSING POLICY LIMITATIONS.....	289
Connecting Assessment Policy to the Purposes of Schooling .....	290
Reducing the Emphasis on Annual Tests.....	292
Educating and Assessing the Whole Learner.....	293
Connecting Outcomes to Inputs Through Opportunity-to-Learn Indicators .....	294
TOWARD ASSESSMENT POLICIES THAT SUPPORT BALANCED ASSESSMENT SYSTEMS .....	296
CONCLUSION .....	301
REFERENCES.....	301

## 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

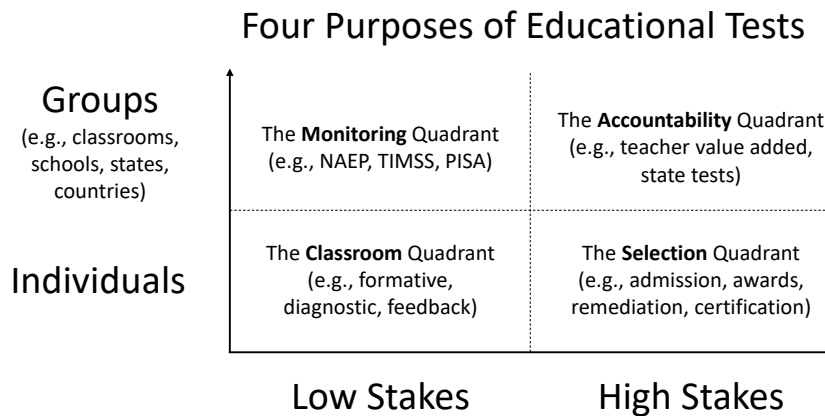
is beyond the scope of this chapter; instead, we focus on *assessment policy*, which we define as policy enacted at the federal, state, or local level that mandates, incentivizes, or supports assessing student learning and other outcomes. This definition incorporates a wide assortment of policies including, 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<sup>1</sup>

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

---

<sup>1</sup> 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; Popho, 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),<sup>2</sup> which originally comprised 24 member states and the District of Columbia, and the Smarter Balanced Assessment Consortium (SBAC),<sup>3</sup> consisting of 15 member states. Additional efforts specifically focused on English learners were launched by the WIDA consortium<sup>4</sup> (originally established by Wisconsin, Delaware, and Arkansas, and now comprising 41 states) and ELPA 21<sup>5</sup> (English Language Proficiency Assessment for the 21st Century, used in 10 states). Finally, the Dynamic Learning Maps Alternate Assessment<sup>6</sup> for students with significant disabilities is administered in 21 member states, and the Multi-State Alternate Assessment<sup>7</sup> 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<sup>8</sup> 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-

---

<sup>2</sup> See <https://resources.newmeridiancorp.org/research>.

<sup>3</sup> See <https://smarterbalanced.org>.

<sup>4</sup> See <https://wida.wisc.edu>.

<sup>5</sup> See <https://www.elpa21.org>.

<sup>6</sup> See <https://dynamiclearningmaps.org/dlm-assessments>.

<sup>7</sup> See <https://www.msaastates.com>.

<sup>8</sup> See <https://www.nj.gov/education/assessment/resources>.

mission of the States<sup>9</sup>). 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).<sup>10</sup>

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),<sup>11</sup> 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.,

---

<sup>9</sup> See <https://www.ecs.org/doctype/50-state-comparison>.

<sup>10</sup> We discuss the SQSS indicator in greater detail in the section titled “Educating and Assessing the Whole Learner.”

<sup>11</sup> 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 (TIMSS).

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](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-IncreasingAccessstoAdvancedCoursework.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-Responsibility\\_v8-single.pdf](https://www.aspeninstitute.org/wp-content/uploads/2022/09/Aspen-Ed-Soc-Opportunity-Responsibility_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://measuringself.caseli.org/wp-content/uploads/2019/09/AWG-State-of-the-Field-Report\\_2019\\_Exec\\_Summary\\_DIGITAL\\_Final-1.pdf](https://measuringself.caseli.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://edsources.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](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](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](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](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](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>.
- Hewlett Foundation. (2013). Deeper learning competencies. [https://hewlett.org/wp-content/uploads/2016/08/Deeper\\_Learning\\_Defined\\_April\\_2013.pdf](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](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](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](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](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](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](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](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](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/michaelt Nietzel/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/FutureEdTestingLandscapeReport.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](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](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., Jofstus, 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., Slegers, 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](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](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](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>.

## Biographical Sketches of Steering Committee Members, Authors, and Reviewers

**Jared Anthony** is the director of the Carnegie Foundation for the Advancement of Teaching Postsecondary Commission, where he is working to accelerate the success of underrepresented students in their secondary and postsecondary education and as they progress toward meaningful careers. Prior to the Carnegie Postsecondary Commission, Anthony served as the director of policy and assessment redesign for the Colorado Education Initiative (CEI), where he helped deepen CEI's understanding of policy and practice related to assessment and school quality measures, as well as coaching schools and districts in implementing strategies for creating balanced assessment systems. Prior to CEI, Anthony served as the assessment principal consultant for the Colorado Department of Education. In this role, he partnered with higher education institutions, districts, and schools to increase student access to higher education using evidence from performance assessments. He also coordinated Colorado's high school accountability assessments and worked to increase assessment literacy among Colorado's districts and schools. Anthony holds a Ph.D. in educational psychology from Fordham University, where he focused on how student motivation and self-regulated learning can influence academic success.

**Courtney Bell** is the director of the Wisconsin Center for Education Research (WCER) and a professor of learning sciences at the University of Wisconsin–Madison. After receiving a postdoctoral fellowship in measurement from the American Educational Research Association and Educational Testing Service (ETS), Bell worked as a researcher at ETS—the world's largest private nonprofit educational testing and assessment organization—for more than a decade. Since taking the helm at WCER, Bell has led the development of two international teacher observation systems and served as a principal investigator on the Global Teaching InSights study—the first of its kind to comprehensively measure teaching quality using observations, artifacts, questionnaires, and student outcomes in eight economies. Bell is passionate about understanding

and improving teaching for historically underserved children. Her interdisciplinary collaborative work is situated at the intersections of research, policy, and practice and spans the issues of parental choice, teaching performance assessments, teaching quality, teacher learning, teacher education, international comparisons of teaching, and measuring teaching. Bell is currently engaged in both national and international studies of teaching, teacher education, and teacher learning. A former high school science teacher, Bell holds a Ph.D. in curriculum, teaching and educational policy from Michigan State University and a B.A. in chemistry from Dartmouth College.

**Amy I. Berman** is the deputy director of the National Academy of Education (NAEd), where she works to advance the organization's strategic and research initiatives. Prior to NAEd, Berman was an education civil rights lawyer, serving as an enforcement director at the U.S. Department of Education's Office for Civil Rights and the section chief at the U.S. Department of Justice, Civil Rights Division, Educational Opportunities Section. In these positions, Berman worked to ensure equal access to education through the vigorous enforcement of civil rights laws, including in the areas of race, national origin, sex, religion, disability, and language. In addition to enforcement, Berman worked on key guidance documents addressing the use of race in schools, harassment in schools, education of English learner students, and the requirement to educate all students regardless of immigration status. She has served as an adjunct professor at The George Washington University Graduate School of Education and Human Development (GSEHD), The George Washington University Law School, and the American University Washington College of Law. She recently co-edited a volume for the NAEd titled *Comparability of Large-Scale Educational Assessments: Issues and Recommendations* (2020) and a volume for The ANNALS of the American Academy of Political and Social Science titled *What Use Is Educational Assessment?* (2019). Berman holds an Ed.S. in education policy from GSEHD, a J.D. from Harvard Law School, and a B.S. in industrial and labor relations from Cornell University.

**Linda Darling-Hammond** is the president and chief executive officer of the Learning Policy Institute, created to provide high-quality research for policies that enable equitable and empowering education for each and every child. She is also the Charles E. Ducommun professor of education emeritus at Stanford University, where she founded the Stanford Center for Opportunity Policy in Education. She also served as a faculty sponsor for the teacher education program at Stanford University, which she helped to redesign. Darling-Hammond is the past president of the American Educational Research Association, a member of the National Academy of Education, and a member of the American Academy of Arts & Sciences. She was recently appointed by President Biden as a member of the National Board of Education Sciences, which oversees the Institute for Education Sciences. In 2022, Darling-Hammond received the Yidan Prize for Education Research in recognition of her work, which has shaped education policy and practice around the most equitable and effective ways to teach and learn. Darling-Hammond is the author or editor of more than 30 books and 600 other publications on teacher quality and educational equity, including *Teaching as the Learning Profession* (1999), *Preparing Teachers for a Changing World: What Teachers Should Learn and Be Able to Do* (2007), *Powerful Teacher Education* (2013), and *Preparing Teachers for Deeper Learning*



(2019). Darling-Hammond holds an Ed.D. from Temple University and a B.A. from Yale University.

**Kathryn Dewsbury-White** is the president and chief executive officer of the Michigan Assessment Consortium (MAC), where she directs the MAC's work in pursuit of increasing assessment literate practice among Michigan's education stakeholders. The MAC, founded in 2008, supports numerous professional learning programs and networked communities, undertakes proof-of-concept evaluations for its assessment programs and services, and publishes resources in formats designed to support practicing educators. The MAC's services, publications, and informal applied research are designed to leverage the implementation of balanced and learner-centric assessment systems. Dewsbury-White also serves as the project or contract director for the MAC's grants and secures and cultivates the many national and state partnerships and collaborations necessary to advance its mission. Prior to the MAC, Dewsbury-White served as a curriculum and professional learning director for a large education service agency for 25 years, directing curriculum and assessment review projects and professional learning programs through the evolution of the standards-based education movement. She initiated and developed a sustaining arts integration professional learning program with the Wharton Center for the Performing Arts and the Kennedy Center for the Performing Arts in mid-Michigan. She also co-founded the Early College at Lansing Community College and served as its first principal. Early in her career, she taught social studies and English language arts to adult and alternative education students. Dewsbury-White holds a Ph.D. in curriculum and instruction, an M.A. in reading, and a B.A. in political philosophy and secondary education from Michigan State University.

**Elena Diaz-Bilello** is the associate director of the Center for Assessment, Design, Research, and Evaluation (CADRE) and a faculty affiliate in the School of Education at the University of Colorado Boulder. Diaz-Bilello collaborates with state education agencies, school districts, and educational organizations to develop practical and sound approaches to address assessment and educational policy challenges. This partnership work includes conducting mixed-methods research and program evaluations to study the effectiveness of educational reforms and initiatives intended to provide equitable learning and assessment experiences, and to improve teaching and learning. She also serves as a Technical Advisory Committee member in several states. Prior to joining CADRE, Diaz-Bilello was a senior associate at the National Center for the Improvement of Educational Assessment providing technical assistance and guidance to state agencies, the U.S. Department of Education, national organizations, and school districts in the areas of designing validity and program evaluation studies and improving upon accountability and assessment practices. Early in her K-12 career, she conducted district-wide program evaluations in the Denver Public Schools and served as an advisor/consultant to the Colorado Department of Education. Diaz-Bilello holds a Ph.D. in research and evaluation methodology from the University of Colorado Boulder, an M.P.A. in international economic and political development from Columbia University, and a B.A. in philosophy and history from Lewis and Clark College.

**Debbie Durrence** is the executive director of data governance for the Gwinnett County Public Schools in metro Atlanta, Georgia, which serves a diverse population of more than 180,000 students. She oversees teams focused on student data management, data integration, data governance, state and federal reporting, data privacy, and data systems and engineering. Prior to serving as executive director of data governance, Durrence served as executive director of accountability and assessment, where she led a team that developed and administered a district assessment program, provided staff development programs for assessment literacy, and supervised the administration of all standardized assessments. In addition to serving as a member of the board of directors and the classroom assessment committee for the National Council on Measurement in Education (NCME), she currently serves on the board of directors for 1EdTech and the K–12 advisory for Qualtrics. She has presented on a variety of topics during conferences for organizations including NCME, the American Educational Research Association, 1EdTech, the National Center for the Improvement of Educational Assessment, and the Ed-Fi Alliance. Durrence holds an Ed.D. in education leadership from the University of Georgia, where her dissertation focused on the effectiveness of technology training in a K–12 setting.

**John Q. Easton** is a senior advisor to the Institute for Policy Research at Northwestern University. Prior to Northwestern, he served as a senior fellow at the University of Chicago Consortium on School Research (the Consortium), as well as the Consortium’s deputy director and executive director. Easton also served as the director of the Institute of Education Sciences and the vice president for programs at The Spencer Foundation. He is involved in several advisory boards at non-profit organizations, including the Illinois Economic Security Advisory Board. Easton is the chair of the advisory boards for both the Illinois Workforce and Education Research Collaborative and the Early Childhood Research Alliance of Chicago, and recently completed a two-year term on the Chicago Public Schools’ Accountability Redesign Advisory Group. Easton was recently appointed to the 2023–2024 EdWorkingPapers Review Board at the Annenberg Institute at Brown University. Easton holds a Ph.D. in education from the University of Chicago, an M.S. in psychology from Western Washington University, and a B.A. in psychology from Hobart College.

**Carla Evans** is a senior associate at the National Center for the Improvement of Educational Assessment. Evans works primarily with state education agencies and other entities on projects that bridge classroom and large-scale assessment. Her research interests and expertise focus on the impact and implementation of assessment and accountability policies on teaching and learning. For example, she conducts policy research related to balanced and innovative assessment systems, culturally responsive assessment, performance-based assessments, and assessment literacy initiatives. Evans has received numerous honors, including the American Educational Research Association’s (AERA’s) Division H Outstanding Dissertation Award, the University of New Hampshire’s (UNH’s) Dissertation Year Fellowship, UNH Graduate Research Assistantships, and the UNH Education Department Outstanding Graduate Student Paper Award. Evans currently serves on AERA’s Classroom Assessment Special Interest Group’s leadership team and as an adjunct professor of education at UNH. Evans

has published numerous articles in peer-reviewed journals and regularly presents her research at AERA, the National Council for Measurement in Education (NCME), and the National Conference on Student Assessment. She has co-authored, with Scott Marion, a forthcoming book on instructionally useful assessment and is co-editing an NCME volume on culturally responsive assessment. Evans began her career as an elementary classroom teacher and taught for almost a decade. She holds a Ph.D. in education with a concentration in assessment, evaluation, and policy from UNH.

**Erin Marie Furtak** is a professor of STEM education at the University of Colorado Boulder. A former high school biology and earth science teacher, Furtak studies how teachers learn through the iterative design and enactment of classroom assessment, and how formative assessment can be a vehicle for more equitable learning. She conducts extensive service to the teaching profession through long-term research and professional development partnerships with school districts and organizations in Colorado and across the United States. Furtak received the 2011 Presidential Early Career Award for Scientists and Engineers and the German Chancellor Fellowship from the Alexander von Humboldt Foundation in 2006. Her research and professional writing have been published in multiple journal articles, research-and practitioner-oriented books, book chapters, humorous essays, and advice columns. Her most recent book, *Formative Assessment for 3D Science Learning: Supporting Ambitious and Equitable Instruction*, was published in 2023 by Teachers College Press. Furtak holds a Ph.D. in curriculum and teacher education from Stanford University; an M.A. in education from the University of Denver; and a B.A. in environmental, population, and organismic biology from the University of Colorado Boulder.

**Susan R. Goldman** is a distinguished professor emerita of liberal arts and sciences, psychology, and education and a founding co-director of the Learning Sciences Research Institute at the University of Illinois Chicago. She has conducted research on subject matter learning, instruction, assessment, and roles for technology throughout her career. Goldman focuses on understanding literacy demands in different academic disciplines and the implications of these demands for supporting subject matter learning. Goldman has also been involved in Project READi throughout her career, and more information about the Project can be found at <https://projectreadi.org>. Her most recent research focuses on how teachers learn to engage in instruction and assessment consistent with the deep learning needed to meet 21st-century demands. This research highlights the importance of systemic support for change, as well as necessary collaboration with teachers and school and district-level practitioners. Goldman is a member of the American Academy of Arts & Sciences and the National Academy of Education, an inaugural fellow of the Society for Text and Discourse and the International Society of the Learning Sciences (ISLS), and a fellow of the American Educational Research Association. She served on the ISLS Board of Directors from 2009–2015, as its president from 2012–2013, and as its executive officer from 2016–2023. She also served on the Board of Directors of the Society for Text and Discourse from 1990–2007 and as its president from 2000–2007. Goldman served as an associate editor for *Cognition and Instruction*, *Discourse Processes*, and *Journal of Educational Psychology*. She currently serves on the editorial boards of the *Journal of the Learning Sciences*, the *International Journal of Computer-Supported Collabora-*

*tive Learning, and Research in the Teaching of English.* Goldman holds a Ph.D. in psychology from the University of Pittsburgh, an M.S. in psychology from the University of Pittsburgh, and a B.A. in psychology from Barnard College of Columbia University.

**Brian Gong** is a senior associate with the National Center for the Improvement of Educational Assessment (the Center for Assessment), where he assists states and other educational entities in addressing challenging assessment and accountability design and implementation issues. Gong has helped develop educationally valuable and technically defensible state accountability systems and innovative assessments. He has also been involved with creating policies, models, and criteria for promoting validity, reliability, and credibility in both assessments and accountability systems through work with groups such as the U.S. Department of Education, the Council of Chief State School Officers (CCSSO), the National Center on Educational Outcomes, and several state Technical Advisory Committees. His recent work has also included the development of standards and criteria for the design and evaluation of assessment and accountability programs. Gong was a member of the committee tasked with revising the Standards for Educational and Psychological Testing and the co-author of content methodology to implement the CCSSO Criteria for Procuring and Evaluating High-Quality Assessments. Prior to co-founding the Center for Assessment in 1998, Gong was responsible for curriculum, assessment, and accountability in the Kentucky Department of Education and served as a research scientist at the Educational Testing Service. Gong holds a Ph.D. in education with a concentration in the design and evaluation of educational programs from Stanford University.

**Ajit Gopalakrishnan** is the chief performance officer for the Connecticut State Department of Education (CSDE), where he oversees data collection, student assessment, psychometrics, reporting, research, and school and district accountability. Prior to CSDE, he worked in the field of adult education for more than 15 years, where he managed initiatives on standards, assessment, data, and accountability. He also served as Connecticut's General Education Development administrator. Gopalakrishnan has published articles on learner retention, accountability, and technology. Gopalakrishnan holds an M.B.A. in management and industrial relations for the University of Bridgeport and a B.B.A. from Loyola College.

**Edward Haertel** is the Jacks Family professor of education, emeritus, at Stanford University. While at Stanford, he studied quantitative research methods, psychometrics, and educational policy—especially test-based accountability and the use of test data for educational program evaluation. Haertel's early work investigated the use of latent class models for item response data, and his later projects included studies of standard setting and standards-based score interpretations, statistical properties of test-based accountability systems, metric-free measures of score gaps and trends, and examining value-added models for teacher evaluation from a psychometric perspective. Prior to his retirement, Haertel served as the president of the National Council on Measurement in Education; the chair of the Technical Advisory Committee concerned with the design and refinements of California's test-based school accountability system; the chair of the National Academies of Sciences, Engineering, and Medicine's Board on Testing and

Assessment; and as a member of the National Assessment Governing Board. Haertel holds a Ph.D. in measurement, evaluation, and statistical analysis from the University of Chicago.

**Laura S. Hamilton** is a senior director at the American Institutes for Research (AIR), where she leads the Center on Advancing Measurement and Assessment. This interdisciplinary center promotes high-quality, innovative approaches to measurement and assessment, generating evidence to inform policy and practice in education, workforce, and family and community development. Hamilton also leads and conducts research on education policy, teaching, educational technology, youth development, and civic learning and engagement. Prior to AIR, Hamilton served as an associate vice president in the Research and Measurement Sciences area at the Educational Testing Service (ETS), leading a portfolio of research on assessment and learning related to K–12 and postsecondary education, workforce development, and language assessment. Prior to ETS, she held the distinguished chair in learning and assessment at the RAND Corporation, where she co-directed the American Educator Panels and directed The RAND Center on Social and Emotional Learning Research. Hamilton has served on numerous editorial teams, committees, and panels, including the National Academies of Sciences, Engineering, and Medicine’s Committee on Developing Indicators of Education Equity; the American Educational Research Association / American Psychological Association / National Council on Measurement in Education’s Committee to revise the *Standards for Educational and Psychological Testing*; the Research Advisory Council for the Collaborative for Academic, Social, and Emotional Learning; the Alliance for Learning Innovation Task Force on Inclusive R&D; and Technical Advisory Committees for several state assessment programs. She has held several editorial roles and received the Joseph E. Zins Distinguished Scholar Award for Social and Emotional Learning Action Research. Hamilton holds a Ph.D. in educational psychology and an M.S. in statistics from Stanford University.

**Margaret Heritage** is an independent education consultant whose career has spanned research and practice. She spent 22 years at the University of California, Los Angeles (UCLA), first serving as the principal of the laboratory school of the Graduate School of Education and Information Studies and then as an assistant director at the Center for Research on Evaluation, Standards and Student Testing. While at UCLA, she also taught in the Teacher Education Program. Prior to UCLA, Heritage served as a senior scientist at WestEd. Prior to working in the United States, she held several roles in her native England: classroom teacher, elementary school principal, adjunct professor in the Department of Education at the University of Warwick, and a county inspector of schools. Her current work centers on formative assessment, including how it supports regulatory processes and contributes to educational equity goals, and how formative assessment practices can support English learners in reaching content and language goals.

**Ethan L. Hutt** is the Gary Stuck faculty scholar in education and an associate professor in the School of Education at the University of North Carolina at Chapel Hill. Hutt’s research is concerned with the systems and tools that assess the work of schools—



specifically the academic standards, data systems, and evaluation metrics that have become ubiquitous in modern school systems—resulting in research focused on the historical development, modern use, and ongoing influence of standardized tests, grades, attendance policies, teacher value-added measures, longitudinal datasets, and accountability systems. Hutt is the co-author, with Jack Schneider, of *Off the Mark: How Grades, Ratings, and Rankings Undermine Learning (But Don't Have To)* (2023); the co-editor, with Michael A. Gottfried, of *Absent from School: Understanding and Addressing Student Absenteeism* (2019); and has published more than 25 peer-reviewed articles in a wide variety of venues including *Social Science History*, *Educational Researcher*, *Journal of Teacher Education*, and the *Virginia Law Review*. Hutt has served as an associate editor for *Educational Researcher*, the book review co-editor of *History of Education Quarterly*, and the co-editor-in-chief of *The High School Journal*. Hutt holds a Ph.D. in education from Stanford University, an M.A. in history from Stanford University, and a B.A. in history from Yale University.

**Erika Landl** is a senior associate with the National Center for the Improvement of Educational Assessment (the Center for Assessment), where she provides technical support to states on issues related to the design, development, implementation, and validation of assessment and accountability systems. Prior to the Center for Assessment, Landl was a senior research scientist at Pearson, where she served as the lead measurement specialist for a variety of state and national assessment programs. During her 13 years at Pearson, Landl was responsible for the planning, management, and coordination of the full array of psychometric activities necessary to sustain a large-scale assessment program, including test design and development, scaling and equating, item and test analysis, parameter estimation, standard setting, the development of reliability and validity research, report design, and the creation of technical documentation. Landl serves on multiple state Technical Advisory Committees and is a co-facilitator of the Council of Chief State School Officers' Balanced Assessment Systems Collaborative. Accordingly, a significant portion of her work is focused on developing tools and processes that inform the selection and use of assessment resources by states, districts, and schools. Landl also frequently develops white papers, publications, and presentations for technical and non-technical audiences. Landl holds a Ph.D. in educational measurement and statistics and an M.A. in educational psychology from the University of Iowa.

**Carol D. Lee** is the Edwina S. Tarry professor of education, emerita, in the School of Education and Social Policy and in African American Studies at Northwestern University. She is the president of the National Academy of Education, a past president of the American Educational Research Association (AERA), a former representative to the World Educational Research Association on behalf of AERA, past vice president of AERA's Division G, past president of the National Conference on Research in Language and Literacy, and past co-chair of the Research Assembly of the National Council of Teachers of English (NCTE). She is a member of the National Academy of Education; a fellow of AERA, the National Conference on Research in Language and Literacy, and the International Society of the Learning Sciences; a former fellow at the Center for Advanced Study in the Behavioral Sciences; and a member of the American Academy

of Arts & Sciences and the Reading Hall of Fame. Lee has received numerous awards, including Distinguished Contributions to Research in Education and Scholars of Color Distinguished Scholar Award from AERA; the Harold W. McGraw, Jr. Prize in Education; the James R. Squire Award and the Distinguished Service Award from NCTE; the LAS Alumni Distinguished Service Award from the College of Liberal Arts & Sciences at the University of Illinois-Urbana; The President's Pacesetters Award from the American Association of Blacks in Higher Education; the Lifetime Achievement Award from the American Association of Colleges for Teacher Education; and an honorary doctorate from the University of Pretoria, South Africa. In 2023, she was selected by President Biden to serve on the National Board for Education Sciences. Lee holds a Ph.D. in education (curriculum and instruction) from the University of Chicago, an M.A. in English from the University of Chicago, and a B.A. in the teaching of secondary school English from the University of Illinois at Urbana-Champaign.

**Peter Leonard** is the executive director of Student Assessment & Multi-Tiered Systems of Support (MTSS) for Chicago Public Schools (CPS). He and his team empower CPS stakeholders with high-quality evidence of student learning to advance achievement, access, and opportunity for all students. They achieve this by leading policy, strategy, implementation, and support across all assessment, MTSS, and high-dosage tutoring programs. Leonard also represents CPS in national- and state-level committees, including as the vice chair of the Illinois State Assessment Review Committee. He is a proud fellowship alum of AmeriCorps, Education Pioneers, the University of Chicago Civic Leadership Academy, and the Erikson Institute's Barbara Bowman Leadership Fellows. Leonard holds a Certificate in Civic Leadership from the University of Chicago, an M.A. in the learning sciences from Northwestern University, and a B.A. in history (honors) from the University of Notre Dame.

**Scott F. Marion** is the executive director of the National Center for the Improvement of Educational Assessment and a national leader in conceptualizing and designing innovative and balanced assessment systems to support instructional and other critical uses. Marion's current projects include designing—and supporting states in implementing—assessment and accountability initiatives; providing technically defensible policy guidance; and implementing high-quality, locally designed performance-based assessments. Marion was recently elected to the National Academy of Education and is one of three measurement specialists on the National Assessment Governing Board, which oversees the National Assessment of Educational Progress. He coordinates and/or serves on 10 state or district Technical Advisory Committees for assessment and accountability. He has served on multiple National Research Council committees, including those that provided guidance for next-generation science assessments, investigated the issues and challenges of incorporating value-added measures in educational accountability systems, and outlined best practices in state assessment systems. Marion is a co-author of the validity chapter in *Educational Measurement* (in press) and a co-author, with Carla Evans, of a forthcoming book on instructionally useful assessment. He has published dozens of articles in peer-reviewed journals and edited volumes, and he regularly presents his work at the conferences of the American Educational Research

Association, the National Council on Measurement in Education, and the Council of Chief State School Officers. Marion holds a Ph.D. in education with a concentration in measurement and evaluation from the University of Colorado Boulder.

**José Felipe Martínez** is a professor of social research methodology at the University of California, Los Angeles (UCLA). Prior to UCLA, he was an associate social/behavioral scientist at the RAND Corporation. His research involves applications of measurement theory and methods to issues in education policy and practice—specifically concerning teacher, school, and program evaluation. Areas of particular focus include systems of multiple measures for teacher and school evaluation and instruments and tools for measuring instructional practice and classroom climate in mathematics and science. His recent projects have involved the development and validation of a range of such measures, including electronic teacher portfolios, student surveys, and observation protocols. His work has been supported by the National Science Foundation, The Spencer Foundation, the William T. Grant Foundation, and the James S. McDonnell Foundation. Martínez teaches courses on measurement, research design, and survey methodology. In 2020, he was the recipient of the American Educational Research Association’s (AERA’s) Palmer O. Johnson Memorial Award for the most outstanding paper published in an AERA journal for his article “Assessing the Assessment: Evidence of Reliability and Validity in the edTPA.” Martínez holds a Ph.D. in education from UCLA; an M.A. from UCLA; and a bachelor’s degree in information science from the Autonomous University of Aguascalientes (Mexico).

**Scott Norton** is a retired educator and most recently served as the deputy executive director of programs at the Council of Chief State School Officers (CCSSO). In this role, he directed and oversaw the development and execution of CCSSO’s programmatic areas, including addressing student expectations, student transitions, teacher workforce, school leadership, and overall school improvement. Prior to CCSSO, Norton was the assessment director and then the assistant superintendent in the Office of Standards, Assessments, and Accountability for the Louisiana Department of Education. He also was a public-school teacher in Louisiana. Norton holds a Ph.D. in educational administration and supervision from Louisiana State University, an M.S. in education administration from Louisiana State University, an M.S. in public administration from Louisiana State University, and a B.S. in education from Tennessee Technological University.

**James W. Pellegrino** is a professor emeritus of liberal arts and sciences, psychology, and education and a founding co-director of the Learning Sciences Research Institute at the University of Illinois Chicago. His research on STEM education and assessment has been funded by the National Science Foundation, the Institute of Education Sciences, and private foundations. His recent projects have focused on the design of high-quality science assessments and instructional resources for K–8 classrooms. He has chaired several National Academies of Sciences, Engineering, and Medicine committees that issued major reports related to education and assessment, including the Committee on the Evaluation of National and State Assessments of Educational Progress; the Committee on Learning Research and Educational Practice; and the Committee on the Foundations of Assessment, which authored the seminal report *Knowing What Students*

*Know: The Science and Design of Educational Assessment*. Most recently, he served on the Committee on Science Learning: Computer Games, Simulations, and Education and the Committee on a Conceptual Framework for New Science Education Standards. He chaired the Committee on Defining Deeper Learning and 21st Century Skills and co-chaired the Committee on Developing Assessments of Science Proficiency in K–12. He is a lifetime member of the National Academy of Education and the American Academy of Arts & Sciences. He currently serves on the National Assessment of Educational Progress Validity Studies Panel and on Technical Advisory Committees for multiple state assessment programs, including Illinois, Maine, New York, Rhode Island, Texas, and Vermont. He has published numerous books, chapters, and articles on cognitive theory and research and their implications for the design of instruction and assessment. Pellegrino holds a Ph.D. and an M.A. in experimental, quantitative psychology from the University of Colorado and a B.A. in psychology from Colgate University.

**William R. Penuel** is a distinguished professor of learning sciences and human development in the Institute of Cognitive Science and School of Education at the University of Colorado Boulder. He designs and studies curriculum materials, assessments, and professional learning experiences for teachers in STEM education—primarily in science. He investigates how contemplative practices and critical inquiry can support educators in cultivating more compassionate learning environments and schools. He also focuses on how long-term research–practice partnerships can be organized to address systemic inequities in education systems linked to race, gender and sexual diversity, and language. In each of his projects, Penuel works in partnership with educators and education leaders to explore how to attenuate inequities in school systems by (1) creating equitable classroom cultures that attend to student experience; (2) testing strategies to address epistemic injustices in whose knowledge is elicited and valued; and (3) connecting teaching to the interests, experiences, and identities of learners, particularly those to whom our society owes an education debt. He uses a wide range of research methods, including one he and his colleagues developed called design-based implementation research, to test what he and his colleagues co-design. Penuel is an author of two books on research–practice partnerships—*Creating Research-Practice Partnerships in Education* (2017) and *Connecting Research and Practice for Educational Improvement: Ethical and Equitable Approaches* (2018) and co-edited a book on improvement research titled *The Foundational Handbook on Improvement Research in Education* (2022). Penuel holds a Ph.D. in developmental psychology from Clark University, an Ed.M. in human development and psychology from Harvard Graduate School of Education, and a B.A. in psychology from Clark University.

**Donald J. Peurach** is a professor of educational policy, leadership, and innovation at the University of Michigan Marsal Family School of Education, a senior fellow at the Carnegie Foundation for the Advancement of Teaching, and a co-director—with Jennifer Lin Russell—of the Improvement Scholars Network. His research focuses on the organization, management, and improvement of instruction in education systems, with a particular focus on network-based continuous improvement. Peurach is the co-editor of *The Foundational Handbook on Improvement Research in Education* (2022); the author of *Seeing Complexity in Public Education: Problems, Possibilities, and Success for*



*All* (2011); and a co-author of *Improvement by Design: The Promise of Better Schools* (2014). Peurach holds a Ph.D. in educational studies from the University of Michigan; an M.P.P. from the University of Michigan; and a B.A. in computer science and his teaching certificate from Wayne State University.

**Morgan S. Polikoff** is a professor of education at the University of Southern California (USC) Rossier School of Education and a co-director of the USC EdPolicy Hub. He researches the design, implementation, and effects of curriculum, standards, accountability, and assessment policies. He has also led state and nationally representative surveys to understand Americans' views on education policy and the educational experiences of American households since the onset of the COVID-19 pandemic. He has been the principal investigator or co-principal investigator on more than \$15 million in federal and foundation grants, including the Institute of Education Science–funded Center on Standards, Alignment, Instruction and Learning from 2015 through 2020. He received the American Educational Research Association (AERA) Early Career Award in 2017 and served on the National Academies of Sciences, Engineering, and Medicine's Committee on Developing Indicators of Educational Equity. Polikoff is a committed public intellectual, disseminating his and others' research through dozens of commentaries, blogs, and social media engagements, winning AERA's Outstanding Public Communication of Education Research Award in 2020 for his impact on state accountability policies under the Every Student Succeeds Act. Polikoff has published more than 50 peer-reviewed journal articles and one book, *Beyond Standards: The Fragmentation of Education Governance and the Promise of Curriculum Reform* (2021). He has served as an associate editor of the *American Educational Research Journal* and a co-editor of *Educational Evaluation and Policy Analysis*. Polikoff holds a Ph.D. in education policy from the University of Pennsylvania and a B.S. in mathematics from the University of Illinois at Urbana-Champaign.

**Jennifer Randall** is the Dunn Family Endowed Professor of Psychometrics and Test Development at the Marsal Family School of Education at the University of Michigan. She is also the founding president of the Center for Measurement Justice, a research center dedicated to facilitating the adoption of justice-oriented educational assessment and measurement processes. Randall began her career as a public high school social studies teacher, working with racially and ethnically minoritized students. In this capacity, she began to recognize the ways that traditional assessment practices cause deep and irreparable harm to the most marginalized students—the students the system should be seeking to serve the most. Her work seeks to disrupt white supremacist, racist logics in assessment through practices that center justice and are explicitly and unapologetically antiracist. Randall sits on numerous state and national working groups, committees, and Technical Advisory Groups as a fierce advocate for antiracist, liberating processes that center the needs of Black, Brown, and Indigenous students. In addition, she teaches courses in measurement theory, statistics, research design, and assessment. Randall's work has been published in assessment and education journals such as *Educational Measurement: Issues and Practice*, *Educational Assessment*, and *Language Testing*. Randall holds a Ph.D. in educational measurement from Emory



University, an M.A.T. in education from Duke University, and bachelor's degree in psychology from Duke University.

**Maria Araceli Ruiz-Primo** is an associate professor in the Graduate School of Education at Stanford University and the director of the Laboratory of Educational Assessment, Research, and InnovatioN. Her work, funded mainly by the National Science Foundation and the Institute of Education Sciences, examines assessment practices and the assessment of student learning in the classroom and large-scale assessment programs. She has researched the development and evaluation of instructionally sensitive assessments and instruments intended to measure teachers' formative assessment practices. She has also recently worked on the analysis of state testing programs. Her publications address the development and evaluation of multiple learning assessment strategies—including concept maps and students' science notebooks—and the study of teachers' informal and formal formative assessment practices, including the use of assessment conversations and embedded assessments. She was a co-editor of a special issue of the *Journal of Research in Science Teaching* on assessment and a special issue of the *Journal of Educational Measurement* on classroom assessment. She has published in *Science, Educational Measurement: Issues and Practice*, the *Journal of Research in Science Teaching*, and other major technical educational research journals.

**Jennifer Lin Russell** is a professor of leadership, policy, and organizations at Vanderbilt University's Peabody College, a senior fellow at the Carnegie Foundation for the Advancement of Teaching, and a co-director—with Donald Peurach—of the Improvement Scholars Network. Her research focuses on organizing educational systems for improvement, with a particular emphasis on networked continuous improvement. Her work seeks to reshape the relationship between educational research and practice to create more equitable learning opportunities for students. She is a co-editor of *The Foundational Handbook on Improvement Research in Education* (2022). Russell holds a Ph.D. in education policy, organizations, measurement, and evaluation from the University of California, Berkeley; an M.A. in curriculum and instruction from the University of San Francisco; and a B.A. in political science and urban studies from Northwestern University.

**Lorrie A. Shepard** is a university distinguished professor in the School of Education at the University of Colorado Boulder. Her research in educational measurement has addressed the use and misuse of tests in education settings. Most cited are her contributions to validity theory, standard setting, bias detection, grade retention, and the effects of high-stakes accountability testing. Her validity research emphasizes the examination of unintended side effects of testing as well as the intended meanings of test constructs. She has conducted validity investigations at the large-scale and classroom assessment level. At the large-scale level, for example, she has served on the National Assessment of Educational Progress Validity Studies Panel (NVS) from 1995 to the present and has conducted studies under the auspices of the NVS. In recent years, her work has focused, at the classroom level, on drawing deeper connections between sociocultural learning theory, ambitious teaching, and formative assessment—with particular attention to culturally responsive pedagogy and equitable assessment practices. Shepard is past

president of the American Educational Research Association (AERA) and past president of the National Council on Measurement in Education (NCME). She is an AERA Fellow, a member of the National Academy of Education (NAEd), and served as the president of NAEd from 2005–2009. Shepard has received distinguished career awards recognizing her contributions in measurement, research, and teacher education, respectively, from NCME and the Educational Testing Service, AERA, and the American Association of Colleges for Teacher Education. Shepard holds a Ph.D. in research and evaluation methodology from the University of Colorado Boulder, an M.A. in counseling from the University of Colorado Boulder, and a B.A. in history from Pomona College.

**Guillermo Solano-Flores** is a professor of education at the Stanford University Graduate School of Education. His research focuses on the intersection of assessment, cultural and linguistic diversity, and fairness. This research is relevant to the testing of students who are not proficient in English in the United States, students from different countries in the context of international comparisons, and students with disabilities. His research is based on the use of multidisciplinary approaches employing a combination of psychometrics, sociolinguistics, semiotics, and cognitive science. He is the author of the theory of test translation error, which addresses testing across cultures and languages. He has also investigated the use of generalizability theory—a psychometric theory of measurement error—in the testing of English language learners and Indigenous populations, and the use of Boolean algebra in complex coding endeavors. Solano-Flores has advised Latin American countries on the development of national assessment systems and has advised countries in Latin America, Asia, Europe, the Middle East, and Northern Africa on the adaptation and translation of performance tasks into multiple languages. His current research projects examine academic language and testing, formative assessment practices for culturally diverse science classrooms, and the design and use of testing accommodations and accessibility resources for students with special needs in computer-administered tests. Solano-Flores serves as a testing and measurement expert for the National Assessment Governing Board and has been a member of Technical Advisory Committees for several assessment projects and institutions, including the Smarter Balanced Assessment Consortium, the National Assessment of Educational Progress, the National Academy of Education, and the Department of Education of Hawaii. Solano-Flores holds a Ph.D. in education with an emphasis in methodology and measurement from the University of California, Santa Barbara; an M.A. in educational psychology from the National University of Mexico; and a B.A. in psychology from the National University of Mexico.

**James P. Spillane** is the Spencer T. and Ann W. Olin Professor in Learning and Organizational Change at the School of Education and Social Policy at Northwestern University. He is also a professor of human development and social policy, a professor of learning sciences, and a faculty associate at Northwestern University's Institute for Policy Research. A former primary school teacher at St. Mary's on the Hill National School in Cork, Ireland, Spillane's work explores the policy implementation process at the state, district, school, and classroom levels, focusing on intergovernmental and policy-practice relations. He also studies organizational leadership and change, conceptualizing organizational leadership as a distributed practice. His recent projects

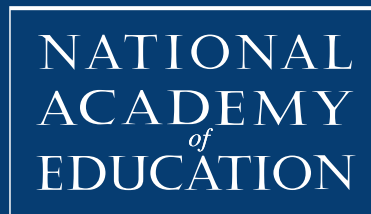
include studies of relations between organizational infrastructure and instructional advice-seeking in schools and the socialization of new school principals. His work has been supported by the National Science Foundation, the Institute of Education Sciences, The Spencer Foundation, The Sherwood Foundation, and the Carnegie Corporation of New York. In recognition of his contribution to educational research, Spillane was elected to the National Academy of Education in 2013 and to the American Academy of Arts & Sciences in 2020. Spillane has published extensively on issues of education policy, policy implementation, school reform, and school leadership. He has authored or edited several books, including *Distributed Leadership* (2006), *Distributed Leadership in Practice* (2007), *Diagnosis and Design for School Improvement: Using a Distributed Perspective to Lead and Manage Change* (2011), *Navigating the Principalship: Key Insights for New and Aspiring School Leaders* (2019), and numerous journal articles and book chapters. Spillane holds a Ph.D. in curriculum, teaching and education policy from Michigan State University and a B.A. in education and geography from St. Patrick's College, National University of Ireland.

**Jonathan A. Supovitz** is a professor of education policy and leadership in the Education Policy Division at the University of Pennsylvania's Graduate School of Education. He is the director of the Consortium for Policy Research in Education, a community of leading education policy scholars from across the United States. Supovitz's research focuses on state and local program and policy designs, implementation, and impacts. His current research focuses on investigating policy innovations that seek to influence the instructional core. He has been a lead and co-investigator of numerous Institute of Education Sciences, National Science Foundation, and other foundation-sponsored research projects and has presented his work nationally and internationally. His latest book, *When Reform Meets Reality: The Power and Pitfalls of Instructional Reform in School Districts* will be released in 2024 by Harvard Education Press. Supovitz holds an Ed.D. in education policy and research methods from Harvard University; an M.A. in public policy from Duke University; and a B.A. in history from the University of California, Berkeley.

**Allison Timberlake** is the deputy superintendent for assessment and accountability at the Georgia Department of Education. She oversees the development, implementation, and evaluation of Georgia's K–12 student assessment and school accountability systems. In her current position, she has led efforts to move Georgia toward a more balanced system of assessments by reducing time spent on state-mandated summative assessments and providing a statewide formative interim assessment resource. She has overseen the development of multiple new assessment systems, including a task-based alternate assessment system for students with significant cognitive disabilities and a game-based formative assessment in literacy and mathematics for first and second grade students. She currently chairs the Council of Chief State School Officers' National Conference on Student Assessment Program Planning Committee, serves as a U.S. Department of Education assessment peer reviewer, and serves as a Technical Advisory Committee member for external organizations. Prior to the Georgia Department of Education, she oversaw assessment and evaluation activities for the Southern Regional Education Board's school improvement initiatives. Timberlake holds a

Ph.D. in educational policy studies—research, measurement, and statistics from Georgia State University and an M.S. and a B.S. in public policy from the Georgia Institute of Technology.

**E. Caroline Wylie** is a senior associate at the National Center for the Improvement of Educational Assessment. Her primary research interests include the design, implementation, and evaluation of balanced assessment systems, with a focus on the use of formative assessment to improve classroom teaching and learning. She has led studies related to the creation of effective, scalable, and sustainable teacher professional development; the formative use of diagnostic questions for classroom-based assessment; assessment literacy; and the role of learning progressions to support formative assessment in mathematics and science. Her research is or has been funded by the Institute of Education Sciences (IES), the National Science Foundation, and the Hewlett Foundation. She is currently the principal investigator of an IES-funded project examining issues of rater quality for formative classroom observations to support peer learning. She serves as a co-advisor for the Council of Chief State School Officers Balanced Assessment Systems Collaborative. Her work has been published in *Applied Measurement in Education*, *Educational Assessment*, *Educational Measurement: Issues and Practice*, *Learning and Instruction*, and *Teachers College Record*, in addition to having contributed to multiple chapters and books. Wylie holds a Ph.D. in assessment and standard setting, a post-graduate certificate in teaching (mathematics), and a B.S. in applied mathematics and physics, all from Queen’s University, Belfast.



The **National Academy of Education** (NAEd) advances high-quality research to improve education policy and practice. Founded in 1965, the NAEd consists of U.S. members and international associates who are elected on the basis of scholarship related to education. The Academy undertakes research studies to address pressing educational issues and administers professional development fellowship programs to enhance the preparation of the next generation of education scholars.

[naeducation.org](http://naeducation.org)