1

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*[1]

CONTENTS

[1] 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).[2] 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-

[2] 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, comprehensiveness, 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[3] (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,

[3] 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[a]

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

[a] 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

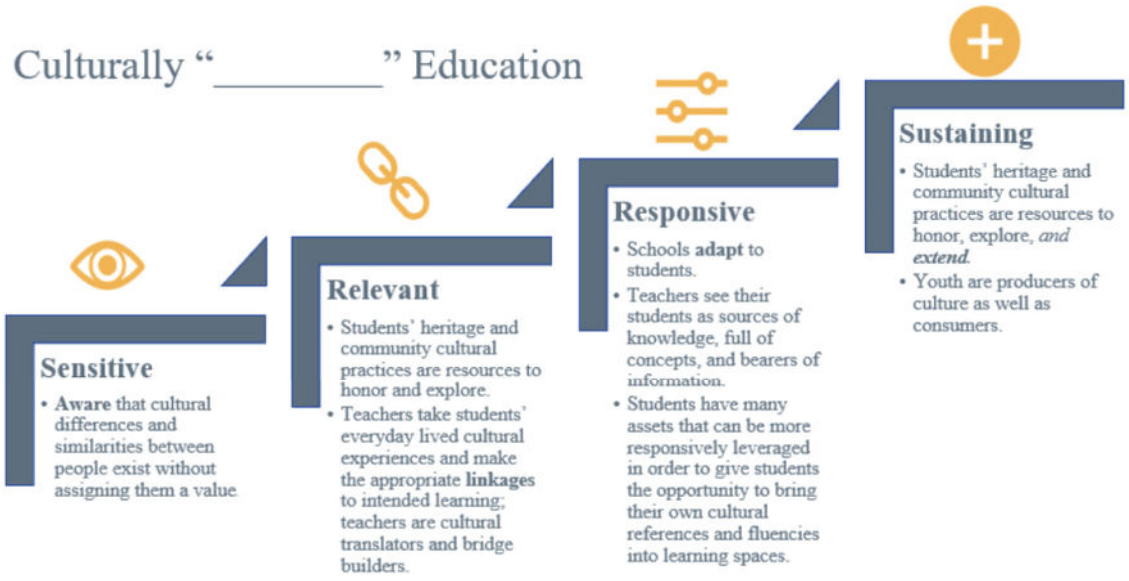

FIGURE 1-1 Culturally "_____" Education.
SOURCE: Evans (2021).

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 sociocultural 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.

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., & Brighouse, 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.