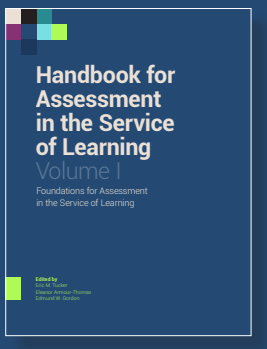
The Cultural Foundations of Learning: Design Considerations for Measurement and Assessment

Roy Pea, Carol D. Lee, Na'ilah Nasir, and Maxine McKinney de Royston

UMassAmherst
University Libraries



Series Editors:
Edmund W. Gordon, Stephen G. Sireci, Eleanor Armour-Thomas, Eva L. Baker, Howard T. Everson, and Eric M. Tucker

ISBN: 978-1-945764-33-2

Suggested Citation:
Pea, R., Lee, C. D., Nasir, N. S., & McKinney de Royston, M. (2025). The cultural foundations of learning: Design considerations for measurement and assessment. In E. M. Tucker, E. Armour-Thomas, & E. W. Gordon (Eds.), *Handbook for assessment in the service of learning, Volume I: Foundations for assessment in the service of learning.* University of Massachusetts Amherst Libraries.

The Cultural Foundations of Learning: Design Considerations for Measurement and Assessment

Roy Pea, Carol D. Lee, Na'ilah Nasir, and Maxine McKinney de Royston

Abstract

In this chapter, we explore the implications of this key insight for the design of assessments. We begin the chapter with an exploration of insights from the most recent science on what learning is and how it happens. We draw from the Science of Learning and Development (SOLD), to explicate key ideas about the nature of learning and the kinds of learning that assessment should serve (Darling-Hammond, Flook, Cook-Harvey, Barron, & Osher, 2020). Then we turn to a discussion of assessment, underscoring that our current system of assessment in the U.S. primarily focuses on sorting, rather than learning (Goldman & Lee, 2024), and in doing so, such assessments too often reify racial and class-based disparities. We then examine how we should be thinking about assessment practices, exploring what might be optimal if we were seeking to assess deep learning. We conclude with a discussion of further implications of this perspective for the development of assessments, including the role of AI, and assessment within disciplines.

At its core, the aim of assessment is to productively understand what and how students learn, with an eye towards improving instruction to support further learning processes and enhance learning outcomes. We argue in this chapter that such an endeavor can only effectively occur if we: 1. Start from what we know about learning; 2. Take as central the social and cultural contexts of young people's lives and learning; and 3. Get clear about the purposes of our assessments.

It is a particularly rich time in education for these discussions (Baird, Andrich, Hopfenbeck, & Stobart, 2017; Baroody & Pellegrino, 2023). All indicators suggest that we have moved past the standards-based movement, with its overemphasis on summative high-stakes assessments and punitive approach to improving instruction and learning outcomes (Darling-Hammond, 2017; Kirst, 2024; Volante, Klinger, & DeLuca, 2024). However, while it is clear what has not worked, it may be less clear what might work for developing assessment and measurement systems rooted in robust theories of learning that can provide insights into the complexities of young people's thinking and learning. It is also unclear what kinds of such systems are feasible given the vast number of constraints in education systems. In this chapter, we explore the implications of taking seriously what we know about learning as we consider the design of assessment and measurement systems.

We begin the chapter with an exploration of insights from the most recent science on what learning is and how it happens. We draw from the Science of Learning and Development (SOLD)—an interdisciplinary science converging from research in the cognitive sciences, psychology, neurosciences, sociology, and other fields—which elucidates the nature of learning in ways transcending centuries-old empiricist notions of learning as a passive and culturally neutral process principally involving individual cognition (Nasir, Lee, Pea, & McKinney de Royston, 2020). We seek to ground the insights in this chapter in some key ideas from this literature about the nature of learning and the kinds of learning that assessment should serve (Darling-Hammond, Flook, Cook-Harvey, Barron, & Osher, 2020). Then we turn to a discussion of assessment, underscoring how our current system of assessment in the U.S. primarily focuses on sorting rather than learning (Goldman & Lee, 2024), and that, in doing so, it too often reifies racial and class-based disparities. The next section takes up the questions of how we should be thinking about assessment practices, exploring what might be possible and/or optimal if we sought to assess deep learning rather than focusing principally on sorting. We turn then to

a discussion of further implications of this perspective for the development of assessments, including the role of AI, and assessment within disciplines.

This chapter highlights several of the design principles for assessment that guide this volume. Specifically, it takes up four of these principles. Our arguments take as core that **assessment equity requires fairness in understanding tasks and adaptation to permit the use of different background knowledge and experience.** Our approach begins with the goal of providing equitable access to high quality teaching and learning and views it necessary for the range of students' background knowledge and experiences, from their homes, communities, and cultures, to be valued for learning to take place. We also call attention to the fact that **assessment design supports the learner's processes, motivation, attention, engagement, effort, and metacognition (self-regulation).** We know from the science of learning and development that to assess learning accurately and with integrity, the assessments themselves need to be designed in ways that works with, rather than against, the ways that learners learn. Thus, considering motivation, engagement, and the learners processes is fundamental to assessing learning well. We also argue that assessments need to be more integrated with learning, which includes the idea that at best, **assessments model the structure of expectations and desired learning over time.** They are learning opportunities in and of themselves, which must reflect the values and the learning process. And finally, our chapter underscores that **feedback from assessment results for learners, teachers, administrators, and families clearly addresses decisions and next steps.** When assessments are more aligned with the learning young people are engaging across the contexts of their lives and reflects the desires of learners themselves and their teachers, families, and communities, then the learning they support will be more effective. But that is not enough—they also must be in communication with those stakeholders to ensure clarity about decisions, implications, and next steps.

What We Know About Learning

A long-time popular conception shared by many educational researchers is that learning is something that happens primarily in the head, involving 'exposure to' and 'uptake of' facts and information. Indeed, the cognitive revolution advanced this perspective by documenting how humans use a variety of active cognitive processes in the accumulation of knowledge and for reasoning about how the world works. This perspective was instantiated in behaviorism, cognitive

processing, and cognitive development models of learning (Greeno, 1998). But the most recent science suggests that learning is a much richer and much more complex process, involving not only cognition, but also processes and systems involving emotion, identity, self-perception, and cultural context (Lee, 2024; Nasir, 2024; Nasir et al., 2020; Osher et al., 2016). Further, these systems that have been conceived of as discrete actually interweave throughout development and are rooted in evolutionary drives that make humans fundamentally social (Cantor & Osher, 2021; Immordino-Yang, 2016; Packer & Cole, 2020). One synthesis of this vast body of interdisciplinary scholarship is provided by the RISE Principles (McKinney de Royston et al., 2020; Nasir et al., 2020).

The RISE Principles offer one effort towards a theory of learning that honors the complex and multi-dimensional nature of learning. The four RISE Principles are that learning is:

1. Rooted in the evolutionary, biological and neurological systems of our bodies and minds, and inseparable from our social and cultural activities;
2. Integrated with all other aspects of development, including cognition, emotion, and the formation of identity—to establish a wide-angle view of the whole child;
3. Shaped by everyday life cultural activities, both in and out of school and across the lifespan; and
4. Experienced in our bodies through coordination with social others and the natural and designed worlds.

The first principle, that learning is *Rooted in the evolutionary, biological and neurological systems of our bodies and minds, and inseparable from our social and cultural activities,* begins at the beginning, with the very nature of our evolutionary biological, and neurological systems. It underscores how humans are fundamentally designed to work in context, in proximity to social others, and to meet our human needs for connection (Immordino-Yang, 2016; Lee, Meltzoff, & Kuhl, 2020; Packer & Cole, 2020). This principle calls our attention to the way that we are 'hard wired' for connection and social interaction. Additionally, while evolution has typically been thought of as the changes in biological systems adapting to circumstances and context overtime, social and cultural systems themselves also adapt in important ways (Packer & Cole, 2020; Turner, 2020). This

intertwining of cultural and biological systems is a core principle of development, and is played out across domains, such as brain development and emotional development (Immordino-Yang, 2016; Lee, Meltzoff, & Kuhl, 2020). Thus, to be human is to learn, because being human requires adaptability in the face of changing contexts.

The second principle, that learning is *Integrated with all other aspects of development, including cognition, emotion, and the formation of identity—to establish a wide-angle view of the whole child,* focuses on the ways in which learning involves integration across developmental domains in a whole person perspective which highlights how emotion, identity, self-perception, and cognition are all brought to bear in the learning process. Further, these processes themselves take place in the specificities of particular social, cultural, and historical contexts, which are in interaction with developmental processes such that these settings offer certain possibilities for development while creating challenges with others. It is through such interactive, situated cultural practices that social, racial/ethnic and gender identities exert influence on our developmental and learning trajectories; as histories of racism and other forms of exclusion challenge some possibilities, including the learners' social conditions, such as poorly resourced schools and fewer opportunities to learners (Darling-Hammond, 2010; Moss et al., 2008). Neuroscience research demonstrates that learning is more effective when learners feel safe and a sense of belonging (Darling-Hammond, 2023, Immordino-Yang, 2016; Steele & Cohn-Vargas, 2013), thereby providing yet another example of how the emotional aspects and social contexts are consequential for learning and integral to learning processes. In this respect, it is worth remembering that Benjamin Bloom's influential taxonomy of psychological domains for education (Bloom & Krathwohl, 1956) and its revised version (Anderson & Krathwohl, 2001) encompassed not only the cognitive and psychomotor manual/physical skills but the affective domain of growth in feelings or emotional areas. It did not treat cognitive developmental processes as dissociable nor in isolation from affective developmental ones.

The third principle, that learning is *Shaped by everyday life cultural activities, both in and out of school and across the lifespan,* focuses squarely on how social context shapes learning. It also emphasizes a point that principle one implied—that learning is ubiquitous, happening all of the time and everywhere, and is thus life-wide and life-deep (Banks, Au, Ball, Bell, Gordon, Gutiérrez, Brice-Heath, Lee, Mahiri, Nasir,

Valdés, & Zhou, 2007). Research has shown that rich and deep learning happens in a variety of settings, from families (Nasir, McKinney de Royston, Barron, Bell, Pea, Stevens, Goldman, 2020; Stevens, 2020), to refugee camps (Brice-Heath, Bellino, & Winn, 2020; Dryden-Peterson, 2016), to games (Pinkard et al., 2017), new media communications (Barron et al., 2013), to local corner stores and basketball courts (Nasir, 2000; Taylor, 2009). Not only does learning happen effectively in a range of settings, but it is noteworthy and even problematic that we tend to privilege the learning that happens in schools and thus undervalue the important knowledge young people bring to schools from their homes and communities (Barron, 2006). Learning also occurs on multiple time scales, and shifts depending on where learners are in the life course (Lee, 2024).

And finally, the fourth principle is that learning is *Experienced in our bodies through coordination with social others and the natural and designed worlds.* This principle highlights that learning involves multiple aspects of ourselves, including our physical selves. Learning, like cognition, occurs throughout our bodies, not only in the brain. Embodiment is central to learning—we learn in and through our bodies (Alibali, 2025; Kontra, Goldin-Meadow, & Beilock, 2012; Nathan, 2021; Shapiro & Spaulding, 2024). This embodiment includes touch or sensorimotor interactions, simulation processes, and kinesthetics, as well as embodied activities more directly connected to communication and expression such as gesture and dance (Vogelstein et al., 2019). The principle of learning as embodied honors that not only is human learning experienced through our bodies, but it does this through social coordinated interactions with others (McDermott & Pea, 2020) and with artifacts (Cole, 1996) created by humans. Bahktin (1981) has argued that even when we are alone, our thinking takes up ideas, beliefs and artifacts created by other human beings. Social forms like language mediate learning—i.e., we learn in and through language—as with tools like symbolic systems and computing and communication technologies (Flores, 2020; Peas, 1994; Rosa, 2016; Vakil, 2024; Valdés, 2004). Symbolic systems embody what Cole calls conceptual artifacts. Indeed, language is a mediator of learning and language is used to position learners into particular identities, whether as members of particular communities and/or as experts or novices in an activity (Green, et. al., 2020). Language is also politicized, wherein some languages or speakers—and their bodies—become stigmatized and get leveraged as rationales for denying them access to learning opportunities. This suggests that our accounts of learning also need to attend to the embodied and

physical aspects of learning, including how we learn kinesthetically and through gesture and other forms of bodily engagement (e.g., Abrahamson & Lindgren, 2014).

Overall, the RISE principles offer a way to view learning that honors learning as an expansive and multi-dimensional process; they also make clear how social and cultural contexts are central to them. If learning is this complex and multi-dimensional and involves so many different interacting and inter-connected developmental domains—what does this mean for assessment? How might we design assessments in the service of learning? We will turn to these questions shortly—first, we consider what assessment is, and what we might be trying to accomplish with it.

What Is Assessment?

Assessments in education tend to focus on academic outcomes in literacy and mathematics, in a way that privileges content knowledge, rather than complex thinking processes (Lee, 2024). The past three decades, in particular, have been an era of standards-based assessments, where students take high stakes tests in math and literacy annually, and the results are used to rank schools, determine student proficiency, and to determine teacher merit pay and school status (Kirst, 2024; Darling-Hammond et al., 2017; Nichols & Berliner, 2007). Research has shown that high stakes assessment creates conditions that undermine learning, such as teachers teaching to the test, cheating, and reducing time in subject matters that are not tested, like science and civics (Darling-Hammond et al, 2017). Further, such assessments reinforce disparities by race and social class and fail to include important knowledge that students bring from their families and communities (Goldman & Lee, 2024). They are aligned with the "dominance of restrictive conceptions of what counts as knowledge in the disciplines and the ontology of the disciplines that are currently restricted to Eurocentric histories of the disciplines." (Lee, 2024, pg. 4). In other words, the assessments we most value in the U.S. fail to capture the breadth of knowledge that constitutes learning. This is, in part, because most assessments used in the U.S. are static, one-time summative measures that are not useful for formative purposes nor to inform shifts in instruction.

This contrast between summative and formative assessment raises a key aspect of the challenge. In his influential paper, philosopher of science Michael Scriven (1967) did not consider formative and summative assessment as two different *types* of evaluation—a frequent misunderstanding. Scriven viewed them as two

different *roles* that evaluation can play. In the formative role for assessments, the teacher or evaluator is taking a constructive approach by emphasizing the input that will help improve a program of instruction and thus improve learning. Whereas in the summative role, the teacher or evaluator is determining the worth of the instructional program by understanding the quality of the learning.

When we are not clear about what assessments are for, and when we are unclear about what they do and do not allow us to understand, we are likely to design assessments that may be practically feasible but conceptually unaligned with our goals. Likewise, when assessments focus on outcomes of learning, rather than including the processes of learning, they are less useful to guide teaching and fall short of being a teaching tool for improving the learning of students.

We know that in their purpose and content, existing standardized tests can be conceptualized as cultural artifacts that are products of our nation's dominant common culture in that they are shaped by it and used for its benefit (Greenfield, 1997; Solano-Flores, 2019). Thus, another central challenge is the way that current forms of assessment, and our cultural assumptions about the purposes of assessment, reinscribe problematic inequalities. This distortion happens in several ways. First, common assumptions about how to do assessment are guided by limited understanding of the nature of learning, as well as aligned with hierarchical and racialized beliefs about learners and learning. Deficit narratives and racial hierarchies undergird many widespread beliefs about assessment (Valencia, 2010). These tendencies make us uncurious about what is happening in schools, and about the patterns of inequity that we see again and again. This is ironic as standardized tests, because of their limited purview of learning, instead are more aptly understood as offering a representation of the policy compliance of a learning system—the classroom, the school, the district—and its relative health, including how it is serving different populations of students.

However, there are few diagnostics about learning processes that can help explain such outcomes that are happening within classrooms, schools, or districts. While some patterns of achievement, engagement and disengagement, or learning disparities can be surfaced by standardized tests across subject matters, how to understand these patterns is not made clearer through these outcomes because there is too much rich and contextual information missing. Thus, such assessments do not leave teachers or parents in a better place to support students,

nor to understand the right next steps for learners. And finally, what we are assessing is really about sorting (Oakes, 2005). This is evident in the ways we use the assessments: to sort students into categories of proficiency, to sort teachers into good and bad, and to sort schools into desirable and undesirable. As a society and as scholars of education, we should hold a bolder, more ambitious vision for assessment. We now turn to a description of some elements of that bolder vision.

How Should We Be Thinking About Assessment In Light of The RISE Principles?

These emerging big ideas around learning and development, captured in the evolving science of learning (SOLD) and summarized in what we call the R.I.S.E. principles, introduce radically new re-conceptualizations that stand in tension with the more siloed conceptions of learning and development that evolved out of the cognitive revolution. We focus on the evolutionary and biological foundations of human learning and development because these foundations urge attention to different aspects of learning and learning environments than traditional conceptions of learning.

Among these big ideas is the proposition that thinking and feelings, the emotional salience we attribute to experience, perceptions of the self, others and settings matter and operate in dialogical relations. Another big idea is that these dialogical relations unfold not in simple linear processes but in contexts of emergence. For example, Fisher, Frey, & Hattie (2016) documented trajectories of individual learners over time and found that the trajectories were not linear, and regressions may often be in service of development (Bever, 1982). The developmental research evolving around dynamic complex systems (Thelen & Smith, 1994) offers theoretical and methodological resources for studying how these dialogic processes of learning and development unfold within and across time.

One guiding idea in studying the complex systems of human learning in sociocultural and material contexts is the continuous mutual influence in real time among components as parts within a whole system. In the case of living organisms like human learners, this means organism–environment relations in an ecosystem (Rogoff, 2023). For example, we can conceive of learners' social interactions in or out of school as an ecosystem in which people act together, in concert, monitoring one another's actions, making next moves that take account of what others are doing. Another guiding notion from ecological system analyses

of the role of human interaction in learning is that participants in interaction use multiple sensory means to monitor one another's actions—not attending to speech alone through hearing but to visually and kinesthetically available information—and they draw on multiple semiotic resources in signaling meaning to one another in the everyday event timescales of microseconds, seconds, and minutes (Goodwin, 2017). Of course there are longer scale timeframes in play as well, as ecological relations exist between cultural practices inherited from our human ancestors are being redeployed in transformed ways to suit the learners' present circumstances (Lemke, 2000; Newell, 1994).

We also know that participation in routine cultural practices, social interactions with others, is central for learning. Further, the artifacts that human communities develop over time also matter for learning. For example, the field of epigenetics has demonstrated both that and how genes follow experience rather than the prior propositions, intellectually rooted in Eugenics, that genes determine human ability and possibility. We know further that among the essential targets of human learning and development from infancy on is what is called social cognition (Carlston et al., 2024), namely an evolving ability to read and respond interactively to the internal states of others. Studies in human infancy have documented both that infants pay more attention to other human beings than to objects and that infants and young children learn through observing and imitating human behaviors (Tomasello, Kruger & Ratner, 1993).

Research grounded in ecological systems theory highlights the significance of time and space in shaping human learning. Specifically, what and how people learn varies across different time scales—ranging from moment-to-moment (microgenetic) learning to developmental changes across the lifespan (Lemke, 2000). This perspective also considers how learning is channeled by the dynamic interplay among various settings and the cultural-historical contexts in which individuals live. For instance, Elder's (2018) longitudinal research on individuals who lived through the Great Depression demonstrates how learning and development across childhood, adolescence, adulthood, and older age were profoundly influenced by the specific resources and constraints framing that era.

Before we address the specific implications for assessments in schools, we should also address how the RISE principles require rethinking the ontology and phenomenology of the academic disciplines we teach within schools.

This rethinking will need to include how knowledge in the academic disciplines is operationalized in everyday contexts, including both the possibilities of connections between everyday knowledge and formal academic disciplines and the differences in such knowledge. Such connections are evident in the field of ethnomathematics (Rosa D'Ambrosio, Orey, Shirley, Alangui, Palhares, & Gavarrete, 2016) and in documented relations between indigenous epistemologies about the natural world and around what we think of as formal science (Medin & Bang, 2014). For example, Indigenous epistemologies concerning the natural world robustly conceptualize the interdependence among humans, animals, plants and other elements of the natural world. This framework stands in contrast to scientific epistemologies in biology that position humans at the top of a hierarchical ladder.

At the same time, there is evolving work in the biology field acknowledging humans as inter- dependently relational with the full breadth of the natural world (e.g., Seymour, 2016), and acknowledging the intertwining of genomes, biomes, microbiomes, and cultural meme pools (Leland et al., 2010). There is a growing understanding that humans are not only not separate from the natural world but are intricately connected and reliant on it for survival, with their actions affecting the environment and vice versa, creating a complex web of interdependencies. Humans rely on the natural world for resources like food, water, and air, and energy while simultaneously influencing the environment through activities like land use changes and energy transformation, and pollution. Ecological impact studies examine how human actions affect ecosystems and the biodiversity within them, considering the interconnectedness of all living organisms. Co-evolution names the idea that humans and the natural world have evolved together, with adaptations on both sides influencing each other over time. Biocultural studies combine biological and cultural factors to understand how humans interact with their environment, including their beliefs, practices, and social structures. Research around narrative sensemaking as an evolutionary disposition of humans (Bruner, 1990) is taken up to capture the diverse pathways through which storytelling (in everyday stories, in formal literature, in music lyrics, in digital media) is taken up across time and across cultural communities.

The point of focusing on the ontologies, phenomenologies, and epistemologies that inform academic reasoning across domains is to inform both the content of what we assess and the dimensions of learning for these domains. It is entirely possible

that an individual may demonstrate epistemological dispositions that are relevant to learning in a domain but not demonstrate adequate content knowledge.

The epistemological and phenomenological dimensions of learning are tools for building conceptual understandings. The ways in which assessments–formative or summative–can provide insights into such multiple dimensions of understanding is important. In some fields like mathematics, we have assessments that will reveal students' conceptual understandings, in part because in formal mathematics (from early to more advanced topics) the creation and manipulation of external representations of how one reasons is foundational to the field. Relative to science, standards such as the Next Generation Science Standards identify these multiple dimensions of scientific knowledge and reasoning, although there are not sufficient assessments available to address all the dimensions of knowledge and reasoning captured in the standards. In the field of literacy–reading, writing, vocabulary within and across disciplines–we do not have such protocols around external representations of processes of reasoning; and we thus tend in our assessments to capture outcomes, for example of comprehension or writing, but not cognitive or material processes of making sense of texts or writing processess. In contrast, when we observe typically on television programs or streaming media like YouTube (DeWitt et al., 2013)–sports programs, cooking programs, arts programs–when experts in the field observe others engaging in its practices, these media almost always de-construct the reasoning behind the individual's or the teams' decision making–making thinking visible in ways instrumental to learning.

What specifically do these big ideas tell us about assessing learning and development in the contexts of schools? What do they tell us about how we might engage in assessments of learning and development that unfold in contexts outside of schools, such as in family life or as people participate in activities within their broader community settings?

The Implications for Assessments

Because learning does not simply unfold inside the minds of individuals but more aptly occurs as described by the RISE principles, assessment systems need to provide windows into the multi-dimensionalities of learning and be ecologically valid. We focus first on systems of assessment, with the understanding that it is not merely what happens in classrooms that contribute to or constrain opportunities to learn. Such assessment systems should include windows into knowledge, the

learner, learning settings, and the organization of learning environments within and across settings (Barron, 2014). This framing builds on Gordon's (2007) notion of intellective competence, which contends that assessments should capture not only declarative knowledge, but also the "ability and disposition to use knowledge, technique, and values…to engage and solve both common and novel problems." Gordon's notion of values, and the importance of applying knowledge to not only familiar but also to novel problems, points to the non-linearity and multi-dimensionality of learning. It also points to the important higher-order ways of thinking that are crucial for adaptability and problem-solving.

Windows into knowledge include attending to and documenting the diversity of ontologies of knowledge, including conceptual, procedural, and/or epistemological forms of knowledge. Systems of assessment based on a more expansive view of learning can also offer insights into who the learner is without making restrictive assumptions about what is or not normative. In particular, these systems of assessments may include items that examine a learner's perceptions of themselves, their competencies, the learning settings they engage in, as well as perceptions related to their own coping, safety, and sense of belonging/connectedness to learning, to a discipline, and/or to a learning environment. Likewise, systems of assessment that are based on a multi-dimensional view of learning will also offer a window into understanding the learning settings in which the learner routinely operates that contribute to their learning. These settings can include family, community, classroom, and school settings, as well as specific policies and practices at district, state, and/or federal levels that can enable or constrain opportunities to learn.

Moreover, systems of assessment based on the RISE principles will be designed in ways that also examine how learning is organized in learning environments within and across settings and the opportunities that do or do not exist within them. Such assessments would include measures that capture opportunities for distributed engagement and exploration, not simply dominated by the teacher or whomever is assumed to have the greatest expertise relevant to the target(s) of learning. Similarly, these systems of assessment would facilitate opportunities for learners to create external representations of reasoning that make their thinking visible and/or allow learners to create and examine multiple modes of representations of reasoning as recommended by the Universal Design for Learning framework for learning materials (Rose, 2000). Finally, systems of assessment that hold the RISE principles

as foundational will recruit learners into accessing and utilizing their everyday repertoires relevant to learning tasks (e.g., language genres and registers; everyday applications of knowledge the learner may have experienced and explored outside of the current learning setting; epistemological orientations, particularly toward complexity and learners' experiences outside of the current setting of learning).

In arguing for ecologically valid assessment systems, we accept the premise that learning unfolds dynamically across settings and recruits multiple resources of the individual and the communities of practice in which the learner(s) engage. Thus, ecologically valid systems both provide windows into the elements of the system, as well as which ones are consequential for facilitating and/or foreclosing learning. For our purposes, crucial elements include the breadth of knowledge and dispositions that influential actors in the system, typically adults, deploy in supporting learning. In our formal education system, these key elements include the knowledge and dispositions, and indeed resources, available to teachers, instructional coaches, and specialized personnel such as social workers, counselors, and administrators at all system levels. Ecologically valid systems of assessment are needed in the United States, for example, as the distribution and quality of such knowledge, dispositions and resources are not equitably distributed. Among other issues, the U.S. is known for its problematic narratives about hierarchies of human communities, restrictive notions of learning as solely cognitive, and for a narrow scope of disciplines taught in school.

For example, LiPing Ma studies elementary school teachers of mathematics in the U.S. and China. Among a cohort of 5th grade teachers, she asked them to solve problems involving division of unlike fractions (Ma, 1999). Teachers from both countries could use the canonical procedure for solving such problems. However, when she asked them why they changed the operator from division to multiplication and inverted the numerator and denominator of the second fraction, not one of the U.S. teachers could explain why. Every Chinese teacher offered multiple conceptual mathematical explanations for why. This challenge involves more than the teacher's conceptual knowledge; equally crucial is what Lee Shulman (1986) described as pedagogical content knowledge—the teacher's understanding of what students need to know and do to engage in sophisticated disciplinary problem solving, the typical difficulties learners face, and the instructional strategies that can support their learning during the knowledge development process.

This difference between the cohort of U.S. and Chinese elementary math teachers is not explained by individual differences, but rather by the systems in place to support robust learning. When asked for her explanation of the differences, Ma explained that in China new teachers are not thrown into the classroom. Rather, new teachers work with master teachers in their school building who collaborate with them in planning, teaching and assessing. This parallels, for example, learning in medicine (Cooke et al., 2010). Graduates from medical school are not simply expected to make diagnoses on their own. Rather they work in long term internships to learn to apply what they learned in theory to practice with real and diverse human beings. We do not have such models of teacher learning either in schools of education or in school districts–with some rare exceptions (e.g., Bank Street: Nager & Shapiro, 2007).

Because we have argued for the breadth of what assessment systems need to provide windows into exploring, this means that those who administer and interpret findings from such assessment windows must have a breadth of knowledge to interpret findings from such assessment tools. These understandings include knowledge of child, adolescent and adult development, knowledge of the multiple dimensions of knowledge construction, and a deep disciplinary knowledge of what is being taught and assessed and how it can be fostered during instruction. These actors include teachers, instructional coaches, counselors, school and district administrators, including members of boards of education. All of these actors do not need the same depth of understanding in each area. For example, the school counselor or member of the board of education does not need deep conceptual understanding of the mathematics being taught by teachers and instructional coaches; but they do need to appreciate the fact that restrictive assessments of procedures and outcomes (such as reading comprehension assessments that only address comprehension outcomes but neglect the cognitive and social processes by which students go about making sense of texts) will not provide them with the kind of consequential knowledge on which to make ecologically valid instructional decisions.

While this agenda may sound overwhelming, there are exemplars of systems of assessment—including systems of assessment for preparing teachers and other actors in the educational system—that encompass the breadth needed to evidence an expansive, multi-dimensional view of learning. We offer the cases of OECD's Programme for International Student Assessment (PISA: Seitzer et al., 2021) and the Japanese Lesson Studies (Lewis et al., 2009).

Exemplar: Program for International Student Assessment (PISA).
We offer the PISA case as an example of how to design a program of assessment that does more than consider outcomes—identifying what works in a system and offering insights as to what needs to be changed. This is particularly important because across large-scale national U.S. data sets like NAEP and international comparisons from PISA and Progress in International Reading Literacy Study (PIRLS), we continue to see socioeconomic status and race/ethnicity associated with historically situated disparities in performance outcomes. We offer PISA as a contrast to the NAEP, which is the only national K–12 educational assessment in the U.S. 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 spans beyond student outcome reporting—it also issues surveys to teachers, administrators, and students. One aim of this broad reach is to document opportunities to learn (e.g., resource allocations, instructional practices), including surveying students about how they perceive each content area. Nonetheless, NAEP surveys are far more limited than those used in PISA because PISA also asks students about their sense of well-being and connections to school. PISA goes beyond cognitive outcomes to attend to social and affective well-being. OECD takes an ecological systems approach to data gathering, analysis, and understanding trends in social disparities around educational equity. In this way, PISA more clearly aligns with the expansive dimensions of learning and development discussed in this chapter.

PISA focuses on group trends over time nationally and, in the case of PISA cross-nationally, as a function of periodic administration to targeted population groups. In this way, PISA captures performance at varying grade and age levels and how those performances change over time. In addition, PISA examines the relationship of these performances to postsecondary outcomes, including participation in higher education and the workforce. PISA does not rely on a single assessment but draws from multiple assessments and surveys to make inferences about longitudinal patterns. These inferences, however, are not about the same populations or sets of students, rather, the large-scale dataset allows for size comparisons across data at different time points in the same participating nations. In this way, the assessments offer an opportunity to infer broad longitudinal trends.

Beyond reporting out about proficiency outcomes, OECD also generates a social disparity report. For example, 2018's PISA Social Disparities report examines how socioeconomic status affects learning outcomes across participating nations and the various factors for these differing outcomes (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 PISA 2018 report also 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). This case illustrates what it takes to develop broad-scale national systems of teaching and assessment that provide the types of deep and wide scope of data that can be analyzed to better understand and explain variation in learning outcomes.

What we see in the PISA data is that the consequences of social background on educational success vary greatly across countries. Results from countries like Estonia, Hong Kong, and Vietnam also demonstrate within-country variability, wherein students who may be presumed to be at risk of failure instead succeed. Across OECD countries, more than one in ten disadvantaged students on average were among the top quarter of achievers in science (op. cit., p. 3). These data also suggest that the poorest students in one region might score higher than the wealthiest students in another country. Extending beyond the learning patterns made visible by the NAEP data, PISA's measures make clear it is not inevitable that disadvantaged students will perform worse than more advantaged students. There are positive contexts in which this result does not occur, invite the study of what Gawande (2007) aptly calls 'positive deviances'. The report concludes with a call for a broader understanding of learning, how learning environments affect learning, and for greater attention to the experiences of disadvantaged students in particular:

> *"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)

Lest we misrepresent PISA as a silver bullet, we must also acknowledge that PISA has been criticized as privileging developing countries and not adequately addressing issues of cultural relevance of content (c.f. Sjøberg, 2016; Teltemann & Klieme, 2017). Even with these critiques, OECD's efforts to address systemic features of educational systems that contribute to PISA outcomes are worth investigating as the data they gather spans far beyond NAEP's current scope.

Exemplar: Japanese Lesson Study.

Broad scale assessments like PISA can be helpful at a high-level to identify and offer insights into that which works in a system and that which needs to be changed. However, for individuals within a system, such as teachers, to learn to navigate that which is working well and that which needs to be changed, there also needs to be systemic support and tools for inquiry. To this end, we offer the example of Lesson Study in Japan—where teachers in school-based communities research their own practices and build this level of investigative and responsive practice into their daily workload and school day (Fernandez & Yoshida, 2004; Lewis et al., 2006).

Lesson study is a form of teacher education widely spread throughout Japan. It was introduced into the U.S. in the late 1990's by education scholars (Stigler & Hiebert, 2009) and was quickly taken up in the early 2000's by mathematics education scholars. While Lesson Study is often woven into pre-service teacher's methodology courses, lesson study has also increasingly been used by in-service

teachers who want to observe, discuss, and improve their pedagogical practices, classroom activities, students' learning experiences, and students' learning outcomes.

There are various forms of Lesson Study. At a top-level, the Lesson Study approach is concerned with how teachers "collaboratively plan, observe, and analyze actual classroom practice" (Lewis, Perry, Hurd & O'Connell, 2006, p 273). Specifically, the purpose of Lesson Study is to construct, through collaboration and observation, a practice-based theory that can be used to study and improve the teaching and learning occurring within a learning environment (Katakami, 2011). The process of Lesson Study can be engaged in and led by students, caregivers, teachers, administrators, and/or scholars, in various combinations. It does not require a top-down approach led by those with the most authority or power within a learning eco-system. Instead, lesson study is a collaborative process that can be engaged in by anyone invested in improving the skills, knowledge, and practices of teachers while also improving the knowledge base of teachers and the teaching profession (Fernandez & Yoshida, 2004).

The process of lesson study is not completely scripted, yet there is a general set of steps we will describe below (Fernandez & Yoshida, 2004). The first step is the collaborative planning of the study lesson. This is an opportunity for those involved to share their ideas about what the lesson should cover, how it should be designed, and what its learning objectives are. This planning draws upon the past teaching and learning experiences of the lesson designers, their understandings and observations of the current group of students who will engage with the lesson, their experience with the curriculum, and so forth. This collaborative process results in a lesson design that will be used to anchor the lesson study. While not explicitly identified in the Lesson Study literature, even this first step of lesson design deeply aligns with the multi-dimensional view of the RISE principles. This alignment can be seen in how Lesson Study takes seriously the social, cultural, emotional, cognitive, and contextual factors of a learning environment, as well as those of the students and teachers in that environment, to be ecologically valid and, ultimately, successful in achieving its learning goals.

The second step in Lesson Study is implementation (aka 'enactment') coupled with observation. This process involves one of the teachers actually teaching the co-developed lesson to their students, while the other members of the collaborative

design group observe the lesson's enactment. The public nature of this enactment requires the observers to also know the lesson well and to use it as a tool to guide their observation, note-taking, and subsequent reflection. In step 3 of the Lesson Study process, the design group reconvenes to reflect and discuss how the lesson unfolded. Each person involved, including the teacher(s) who taught the lesson, shares their observations and reactions to watching and/or engaging in the lesson. This review includes making suggestions about how the lesson could be improved vis a vis how it was implemented and experienced, how opportunities for learning were or were not presented and to whom, how it did or did not achieve the expected learning engagements and/or goals, and so on.

The fourth and fifth steps of Lesson Study involve revision and reteaching. While some design groups may decide to end their work at the third step and allow individual teachers to take it from there, design groups may choose to continue to learn together by building off of the reflections and suggestions to create (step 4) and teach (step 5) an updated version of the lesson design. The re-teaching of the new version of the lesson (step 5), mirrors that of step 2 wherein one group member, presumably a teacher, will again teach the lesson while their colleagues observe. Given the organization and nature of schools, it is unlikely that the same teacher will teach the same lesson to the same group of students a second or a third time. Instead, there is some variability with how the re-teaching occurs and who does it. This is an important aspect of Lesson Study, because the point is not to perfect a lesson nor to study the outcomes on a particular group of students to measure the success of a lesson, because the group values giving as many teachers as possible an opportunity to practice teaching the lesson and cultivating a broader base of experiences from which to learn and grow. These communitarian values are consistent with Lesson Study's purpose—to develop a practice-based theory relevant for studying and improving one another's teaching, as well as improving the learning and learning experiences of students.

The final and last step of Lesson Study, should the lesson be retaught, is to again share observations, reflections, and suggestions about the updated lesson version. As with all steps of the Lesson Study process, especially step 4, it is imperative that detailed notes are taken to document the discussion, the ideas generated, and the decisions, including their rationales. Such documentation offers a useful record for later reference should additional revisions and discussions accrue. It also is a necessary record should the teachers decide to report out or share insights about

their collaboration, their revisions, their enactments, etc. with others, whether at their school or in other professional settings concerned with improving teaching and learning.

These two exemplars—Japanese Lesson Study and PISA—suggest several key conditions that must be met for assessment systems to best support robust learning and teaching:

- Assessment developers need to understand human development, cultural communities, disciplinary knowledges, etc. to create effective assessments.
- Assessments need to include longitudinal and culturally-situated assessments of learners. Such assessments need to leverage multiple modalities, not simply texts.
- Assessments need to include social-emotional learning processes and outcomes; and finally,
- Assessments must examine opportunities for learning within learning environments and not only learners.

We have focused here on systems of assessment, yet there are also exemplars of specific assessment tools that individually address dimensions of learning we have identified. These tools are typically not widely distributed nor used. We cannot do justice to them within this chapter and instead invite those interested in these tools to look at those instances explored and discussed by Goldman and Lee (2024).

We will end this chapter with a discussion of some additional factors and tools that are worth considering in light of the current socio-technical context that influences education and many other societal sectors. In particular, the next section explores the potential that emerging technologies, such as generative AI, have for supporting the development of assessment tools that align with the RISE Principles of learning.

The Prospects of Generative-AI Augmented Assessment

In the emerging socio-technological universe, there is ample enthusiasm for the prospective roles of Generative AI in education's future. Generative AI developments based on large language models (LLMs) are proceeding at an unprecedentedly accelerative pace, affecting virtually every sector that produces learning media artifacts such as alphanumeric text, images, sounds, videos, as part of its information and knowledge production processes—which can now serve as input to large language and image models used to empower further generative AI advances. Researchers (Bick, Blanding, and Deming, 2024) and leading technologists and historians are comparing generative AI to the printing press and other epochal innovations like the World Wide Web.

Accordingly, entire industries are in dramatic transformational states with new companies further developing or exploiting generative AI technologies being funded with tens of billions of dollars. The future of human work itself is in question, which implicates some of the many purposes of education. Increasingly, the education sector is also attracting startup funding for generative AI applications and this trend will likely continue. While we cannot treat these opportunities and risks in any detail here, we can nonetheless point in the directions of inquiry which we anticipate will become promising for those enlisting Generative AI tools to augment our traditional socio-technical practices of K–12 learner assessments. We will begin with a brief preamble on how the RISE principles generally relate to the integration of Generative AI into K–12 assessments before describing three central reasons that Generative AI will be integral to the future of assessment.

As we've argued in this chapter, the RISE principles offer a way to view learning that honors its expansiveness and multiplicity and makes clear how social and cultural contexts are central to the learning process. Yet, how might these principles, and the research undergirding them, relate to the emerging uses of Generative AI for assessing learning? First, generative AI applications employed for assessment of learning congruent with the RISE Principles would need to take account of the expansive conditions of learning, as compared to today's more limited assessment paradigm of *assessing students inside school classrooms and class times* using standard assessments *that are either paper-based or computer-based* in their administration.

The understanding of learning being *Rooted* in the evolutionary, biological and neurological systems of our bodies and minds, and inseparable from our social and cultural activities, suggests that for emerging AI-augmented assessments to be effective, they would need to accompany a learner as they engage in and navigate a variety of socio-cultural activities no matter where they occur. Because learning is *Integrated* with all other aspects of development, including cognition, emotion, and the formation of one's multifaceted identity, to establish a wide-angle view of the whole child, emerging AI-augmented assessments will need to encompass these diverse fields of human functioning and not be restricted to the cognitive domain as they largely are today. This expansion would include such assessments being able to engage a learner's social and emotional states such as safety and belonging.

Due to learning being *Shaped* by everyday life cultural activities, both in and out of school, emerging AI-augmented assessments would be persistent throughout the daily activity rounds of the learner, not only in but outside of school contexts. Moreover, learning is *Shaped* by these activities across the lifespan, therefore assessments could not be one-offs that occur at only one given point for a very limited time, rather they would need to be longer-term and longitudinal to some degree (e.g., across a series of days, weeks, months, years, etc.). Lastly, since learning is *Experienced* in our bodies through coordination with social others and the natural and designed worlds, emerging AI-augmented assessments would more thoroughly engage with the physical bodies, gestures, and social and environmental interactions of learners as they coordinate their ongoing activities inside and out of school, including online virtual spaces. Biometric sensing data and multimedia records of learning interactions during activities in and out of school are likely sources of new assessment-related evidence for learning processes and outcomes, which will require satisfying all appropriate data privacy policies and safeguards.

Three central reasons provide further grounds to believe that Generative AI will be integral to the future of assessment congruent with the RISE principles of learning.[1] The first reason is that thoughtful uses of emerging and rapidly developing generative AI approaches *can yield improved assessment for use across traditional and new purposes.* These emerging approaches illustrate prospects for a dramatic

1 This section builds on text notes co-created by Eva Baker and Roy Pea for their work co-chairing the AI in Education Planning Committee of the National Academy of Education during 2024–2025.

change in the type and complexity of assessment tasks and modalities, because they transition the education research field from easily scorable multiple-choice and survey approaches toward tasks more closely approximating the situations, settings, and motivations involved in more complex learning environments. Examples include extended performance tasks with multiple dimensions, to include estimates and remedies for differences in prior knowledge (which is a persistent equity concern) and a range of cognitive demands, e.g., analysis, problem-solving (Anderson & Krathwohl, 2001; Mayer, 2009).

Recall that assessment practices often build upon Benjamin Bloom's cognitive taxonomy of educational goals and objectives (Bloom & Krathwohl, 1956), which proposed a pyramidal series of increasingly complex fundamental cognitive functions: remembering, understanding, applying, analyzing, evaluating, and creating knowledge. Further improvement likelihoods with AI are enhancing the assessment's relevance to the assessed individual's cultural and other experiential background details and preferences (Bailey & Duran, 2019; Bennett, 2023; Duran, 2020) and metacognitive processes (Fisher, Frey & Hattie, 2016; Yaden et al., in press) including components such as attention (Schwartz & Plass, 2020), motivation, and self-efficacy (Rueda, 2013; Rueda, O'Neil & Son, 2016) that can be readily monitored during the process of assessment and learning with learners' uses of technology-based systems. Congruent with the RISE principles emphasis on lifewide, lifelong and life-deep learning, AI is likely to enable greater empirical attention to affirmative development (Gordon and Bridglall, 2006), following individuals' intellectual growth over time within and across subject matter domains, inside and out of school, to ascertain patterns that can be supported or interrupted to produce outcomes of most value. At present, evidence of the developmental performance of individuals over time and contexts is rarely accessible. Indicators of development could very well be integrated across types and goals of various assessments, rather than limited to a particular type of task or format.

The second reason why Generative AI will be integral to the future of assessment is that *learning will be central to all emerging forms of assessment, when such assessments are re-situated as an integral aspect of learning,* rather than separated from it (Gordon Commission on the Future of Assessment, 2013). This orientation extends far beyond usual views on formative evaluation, which depend upon action taken from learning the results of tests or assessments. Instead, new affordances by large language models using artificial intelligence

(AI) can allow the learning and assessment activities *to be functionally blended from the learner's perspective,* while differentiating them in ongoing analyses. To take one example, simulations and games currently exhibit modestly blended elements (albeit highly structured ones: Schwartz & Plass, 2020; Shute et al., 2021). Learning activities with real-life problems based on domain knowledge will permit continuous and seamless feedback, adaptation, and learner adjustment in individual or collaborative settings, especially when accompanied by mobile platforms such as smartphones for the learners' activities which are contributing to their learning and to their being assessed.

It is very likely that education policy will continue to seek out or require summary documented evidence of learner growth or learning program and policy effects. AI will help enable the acquisition of such data to be sampled from individuals' ongoing learning in contrast to current, distinct, and more ceremonial assessments. Clear attention to learning will require tools for use by learning platform purveyors as well as smart support for teachers' own classroom assessments. AI can provide ways to integrate assessments from disparate sources that use common elements in their designs. Approaches to data collection, reporting, and validity and quality will also undergo substantial change. These modifications will require planning and systematic judgment to assure that AI and assessment together support the growth and success of all students.

Third, generative AI *can address barriers to equity found in current approaches to assessment and learning by addressing learner diversity in new substantive ways.* Equity should be a featured goal of AI and assessment, by supporting the development of insights into the full spectrum of the multitudinous ways in which individuals differ in their interests, knowledge, skills, dispositions, affect, motivations, and other learning-relevant characteristics such as bodily engagement. The goal is to capture patterns of performance, to supply needed background and more targeted prior knowledge or resources, and to incorporate appropriate information about learner preferences, experiences, and aspirations at the individual level for their more consequential educational support than is common today. Computationally mediated learning environments including interactive texts, symbols, graphics, audio, video, and animations will take adaptation far beyond current capabilities. Including contextual cultural elements in task structure along with desirable lexical cues can make learning and

assessment situations both comfortable and challenging for diverse learners. How such options should be developed and deployed will be a continuing scientific, policy, and values conversation and played out in the establishment of federal and state research and development priorities amidst a changing landscape of science and educational policies.

To be clear, AI already exists in assessments, particularly in writing (e.g., Ke & Ng, 2019). However, the present state of the art in AI and formal assessment and testing (e.g., OECD, ETS, Duolingo) attends to the use of AI mostly to improve task or item generation and scoring, yet for the most part uses existing measurement models that center on scalability. Importantly, changes in assessment scope and depth congruent with the RISE learning principles as sketched out above will necessitate the development of new approaches to common quality expectations for assessments. How will validity be ascertained for developmental performance once the metrics combine cognitive, affective, and domain-focused performances across learning contexts? How will reliability be reconceptualized when multiple items are not part of the assessment regimen? Will it be feasible to document assessment status for individual and policy use by using agents such as simulated students to reduce the response burden, time, cost, and delay associated with prototype testing and aggregated results for new assessments? It is certain that the infrastructure used to create learning and assessment designs and implementations will be upended. It will be vital to ensure that the clearest scientific knowledge and best practices in the learning sciences will be used to undergird these new AI-augmented learning and assessment systems.

However, such systems are not without their potential downsides. First, we know that AI systems are only as good as the data inputs they are trained on (Ferrara, 2023). Second, and relatedly, Generative AI systems have the potential to reproduce biases and deficit-oriented conceptions that exist in the broader society (Capraro et al., 2024). Hence, attending to the data inputs and the training of AI systems would need to be done carefully and thoughtfully, holding the RISE principles of learning and the ethical and equity-oriented cautions in mind.

Conclusion

We have argued in this chapter that to improve teaching and learning, and support robust learning, assessment systems in the US require a rethinking and revamping. We have offered one frame to support us in this redesign guided by the RISE principles of learning. The first principle argues that learning is *Rooted in the evolutionary, biological, and neurological systems of our bodies and minds, and inseparable from our social and cultural activities.* This suggests that assessment systems need to better account for the ways in which humans are, by definition, pack animals, who are motivated and learn in social contexts with others around them. The second principle is that *learning is integrated with all other aspects of development, including cognition, emotion, and the formation of identity—requiring for resolution a wide-angle view of the whole child.* This principle helps us hold at center how assessments and assessment systems need to account for issues of motivation, emotion, and connection, and not assume that one can separate the cognitive from other central developmental process. It also means that assessments must themselves be designed as experiences that motivate and engage. The third principle is that *Learning is shaped by everyday life cultural activities, both in and out of school and across the lifespan.* This reminds us that we must be vigilantly expansive in the ways that we design assessment systems, and that we should remember that schools are not the only, and perhaps even not the best) contexts within which learners gain important new knowledge and understandings. Ideally, we could draw on assessment systems to capture what schools tend to miss, and to provide points of leverage for supporting the integration of learning across learning settings as well as to include longitudinal dimensions to track learners over time. And finally, the fourth principle argues that learning is experienced in our bodies through coordination with social others and the natural and designed worlds. Thus, assessment systems, too, must draw on multimodal ways of learning and expression.

These pivots in how we design systems of assessment are critical if we are to move beyond the sorting function of assessment to build systems of assessment that are culturally inclusive and learner-centered, and which will provide important information about learner trajectories that can guide teaching in the future. In other words, we can most productively transform assessment when we build on the expansive understandings of learning that the sciences of learning and development make apparent. When we transform our understanding of learning,

and how to cultivate that learning, then the transformation of assessment systems should follow in alignment.

Importantly, at the heart of any assessment system must be a deep respect for the complexity of the learning process itself, and the intertwining of learning with a wide range of developmental processes and domains. We have also explored the possibility of AI augmented assessments which provide new opportunities to lean into some of these important properties of learning. It is a critical time in the field for these discussions, and perhaps more evident than ever that not only is it imperative that we develop more useful and more robust assessment systems, but that we utilize new tools and technologies based upon more robust understandings of learning to do so.

References

Abrahamson, D., & Lindgren, R. (2014). Embodiment and embodied design. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 358–376). NY: Cambridge University Press.

Alibali, M. W. (2025). Action, attention, and multimodal scaffolding. In Edwards, L.D., & Krause, C.M. (Eds.), *The Body in Mathematics: Theoretical and Methodological Lenses.* Boston: Brill.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Anderson, L. W., & Krathwohl, D. R., et al (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives.* New York: Longman.

Bailey, A., & Duran, R. (2019). Language in practice: A mediator of valid interpretations of information generated by classroom assessments among linguistically and culturally diverse students. In S. Brookhart & J. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 46–62). New York: Routledge.

Baird, J. A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: fields apart? *Assessment in Education: Principles, Policy & Practice, 24*(3), 317–350. https://doi.org/10.1080/0969594X.2017.1319337

Bakhtin, M. M. (1981). *Four essays by M. M. Bakhtin.* In M. Holquist (Ed.). Austin: University of Texas Press.

Banks, J., Au, K., Ball, A.F., Bell, P., Gordon, E., Gutiérrez, K., Brice-Heath, S., Lee, C.D., Mahari, J., Nasir, N.S.,Valdés, G., & Zhou, M. (2007). *Learning in and out of school in diverse environments: Life-long, life-wide, life-deep.* Seattle, WA: LIFE Center.

Baroody, A. J., & Pellegrino, J. W. (2023). Assessment for learning. *Classroom-Based STEM Assessment*, 38.

Barron, B. (2006). Interest and self-sustained learning as catalysts of development: A learning ecology perspective. *Human development, 49*(4), 193–224.

Barron, B. (2014). Formative assessment for STEM learning ecosystems: Biographical approaches as a resource for research and practice. *National Research Council Committee on Out-of-School Time STEM.* Washington, DC: National Research Council.

Barron, B., Gómez, K., Pinkard, N., & Martin, C. K. (2013). *The digital youth network: Cultivating new media citizenship in urban communities.* Cambridge, MA: MIT Press.

Bennett, R. E. (2023). Toward a Theory of Socioculturally Responsive Assessment. *Educational Assessment, 28*(2), 83–104. https://doi.org/10.1080/10627197.2023.2202312

Bever, T. G. (1982). *Regression in the service of development.* In Bever, T.G., et al. (Eds.), Regressions in Child Development (pp. 153–188). NY: Routledge.

Bick, A., Blandin, A., & Deming, D.J. (2024). *The Rapid Adoption of Generative AI.* http://dx.doi.org/10.2139/ssrn.4965142

Bloom, B.S. & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: the classification of educational goals, by a committee of college and university examiners.* Handbook I: Cognitive Domain. New York, NY: Longmans, Green.

Bruner, J. (1990). *Acts of meaning.* Cambridge, MA: Harvard University Press.

Cantor, P., & Osher, D. (Eds.). (2021). *The science of learning and development: Enhancing the lives of all young people.* NY: Routledge.

Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., & Viale, R. (2024). The impact of generative artificial intelligence on socioeconomic inequalities and policy making. *PNAS nexus, 3*(6).

Carlston, D. E., Hugenberg, K., & Johnson, K. L. (2024). Principles of social cognition: historical origins and current status. In D.E. Carlston, K. Hugenberg, & K.L. Johnson (Eds.), *The Oxford Handbook of Social Cognition, Second Edition* (pp. 3–19). Oxford: Oxford University Press.

Cole, M. (1996). *Cultural psychology: A once and future discipline.* Cambridge, MA: The Belknap Press of Harvard University Press.

Cooke, M., Irby, D.M., O'Brien, B.C. (2010). (Eds.), *Educating physicians: A call for reform of medical school and residency.* San Francisco: Jossey-Bass, 2010.

Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future.* NY: Teachers College Press.

Darling-Hammond, L. (2017). Developing and measuring higher order Skills: models for state performance assessment systems. *Council of Chief State School Officers.*

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.

Darling-Hammond, L., Herman, J., Pellegrino, J., Abedi, J., Aber, 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, A., Shulman, L., & Steele, C. M. (2013). *Criteria for high-quality assessment.* Stanford Center for Opportunity Policy in Education, 2, 171–192. Stanford, CA: Stanford University.

Darling-Hammond, S. (2023). *Fostering belonging, transforming schools: The impact of restorative practices.* Palo Alto, CA: Learning Policy Institute.

DeWitt, D., Alias, N., Siraj, S., Yaakub, M. Y., Ayob, J., & Ishak, R. (2013). The potential of YouTube for teaching and learning in the performing arts. *Procedia-Social and Behavioral Sciences, 103,* 1118–1126.

Dryden-Peterson, S. (2016). Refugee education: The crossroads of globalization. *Educational researcher, 45*(9), 473–482.

Duran, R. (2020). Orienting Assessment toward Serving Students Reaction to Contributions of Edmund Gordon: We Need to Change Educational Practice and Assessment Together. *Educational Measurement: Issues and Practice.* 39. https://doi.org/10.1111/emip.12371

Elder, G. H. (2018). *Children of the great depression.* NY: Routledge.

Ferrara, E. (2023). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci, 6*(1), 3.

Fisher, D., Frey, N., & Hattie, J. (2016). *Visible learning for literacy, grades K–12: Implementing the practices that work best to accelerate student learning.* Corwin Press.

Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence,* 6, 1162454.

Flores, N. (2020). From academic language to language architecture: Challenging raciolinguistic ideologies in research and practice. *Theory into practice, 59*(1), 22–31.

Gawande, A. (2007). *Better: A surgeon's notes on performance.* New York: Metropolitan.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in cognitive sciences, 3*(11), 419–429.

Goldman, S. R., & Lee, C. D. (2024). Human learning and development: Theoretical perspectives to inform assessment systems. In Marion, Pellegrino, & Berman (Eds.), *Reimagining Balanced Assessment Systems.* DC: National Academy of Education.

Goodwin, C. (2017). *Co-operative action.* New York, NY: Cambridge University Press.

Gordon, E. W., & Bridglall, B. L. (2006). The affirmative development of academic ability: In pursuit of social justice. *Teachers College Record, 108*(14), 58–70.

Gordon, E. W., McGill, D., Iceman, S., Kelley M., Kalinich, Pellegrino, J. W., & Chatterji, M. (2014), Bringing formative classroom assessment to schools and making it count, *Quality Assurance in Education, 22*(4), 339–352.

Greene, J.L., Brock, C., Baker, W.D., & Harris, P. (2020). Positioning theory and discourse analysis: An explanatory theory and analytic lens. In Nasir, N. I. S., Lee, C. D., Pea, R., & McKinney de Royston, M. (2020). *Handbook of the cultural foundations of learning* (pp. 119–140). NY: Routledge.

Greenfield, P. M. (1997). You can't take it with you: Why ability assessments don't cross cultures. *The American Psychologist, 52*(10), 1115–1124.

Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist, 53*(1), 5–26.

Gutiérrez, K. D., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher, 32*(5), 19–25.

Heath, S.B, Bellino, M.J., & Winn, M. (2020). Adaptive learning across the life span. In Nasir, N. I. S., Lee, C. D., Pea, R., & McKinney de Royston, M. (Eds.), *Handbook of the cultural foundations of learning* (pp. 247–260). NY: Routledge.

Immordino-Yang, M. H. (2016). Emotion, sociality, and the brain's default mode network: Insights for educational practice and policy. P*olicy Insights from the Behavioral and Brain Sciences, 3*(2), 211–219.

Ke, Z., & Ng, V. (2019). Automated essay scoring: A survey of the state of the art. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 6300–6308.

Kirst, M. W. (2024). *Standards-based education reforms: Looking back to look forward.* Palo Alto, CA: Learning Policy Institute. https://doi.org/10.54300/647.623

Kontra, C., Goldin-Meadow, S., & Beilock, S. L. (2012). *Embodied learning across the life span. Topics in cognitive science, 4*(4), 731–739.

Laland, K. N., Odling-Smee, J., & Myles, S. (2010). How culture shaped the human genome: bringing genetics and the human sciences together. *Nature Reviews Genetics, 11*(2), 137–148.

Lee, C. D. (2024). Implications of the Science of Learning and Development (SoLD) for Assessments in Education. In R.E. Bennett, L. Darling-Hammond, & Badrinarayan, A. (Eds.), *Socioculturally Responsive Assessment: Implications for Theory, Measurement, and Systems-Level Policy.* Routledge

Lee, C. D., Meltzoff, A. N., & Kuhl, P. K. (2020). The braid of human learning and development: Neuro-physiological processes and participation in cultural practices. In Nasir, N., Lee, C., Pea, R., & McKinney de Royston, M. (Eds.), *Handbook of the cultural foundations of learning* (pp. 24–43). NY: Routledge.

Lemke, J. L. (2000). Across the scales of time: Artifacts, activities, and meanings in ecosocial systems. *Mind, culture, and activity, 7*(4), 273–290.

Lewis, C. C., Perry, R. R., & Hurd, J. (2009). Improving mathematics instruction through lesson study: A theoretical model and North American case. *Journal of Mathematics Teacher Education,* 12, 285–304.

Ma, L. (1999). *Knowing and teaching elementary mathematics.* Mahwah, NJ: Erlbaum.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge: Cambridge University Press.

McDermott, R., & Pea, R. (2020). Learning "how to mean": Embodiment in cultural practices. In Nasir, N., Lee, C., Pea, R., & McKinney de Royston, M. (Eds.), *Handbook of the cultural foundations of learning* (pp. 99–118). NY: Routledge.

McKinney de Royston, M., Lee, C., Nasir, N. I. S., & Pea, R. (2020). Rethinking schools, rethinking learning. *Phi Delta Kappan, 102*(3), 8–13.

McNeill, D. (2016). *Why we gesture: The surprising role of hand movements in communication.* NY: Cambridge University Press.

Medin, D. L., & Bang, M. (2014). *Who's asking? Native science, western science, and science education.* Cambridge, MA: MIT Press.

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *The American Psychologist, 30*(10), 955–966.

Mislevy, R. (2018). *Sociocognitive foundations of educational measurement.* NY: Routledge.

Moss, P. A., Pullin, D. C., Gee, J. P., Haertel, E. H., & Young, L. J. (Eds.). (2008). *Assessment, equity, and opportunity to learn.* NY: Cambridge University Press.

Murphy, R. F. (2019). Artificial intelligence applications to support K–12 teachers and teaching. *Rand Corporation, 10.*

Nager, N., & Shapiro, E. (2007). A progressive approach to the education of teachers: Some principles from Bank Street College of Education. *Occasional Paper Series*, *2007*(18), 1.

Nasir, N. S. (2000). "Points ain't everything": Emergent goals and average and percent understandings in the play of basketball among African American students. *Anthropology & Education Quarterly*, *31*(3), 283–305.

Nasir, N. I. S. (2024). A vision for the future of learning. *Educational Researcher*, 0013189X231222223.

Nasir, N. S., Lee, C. D., Pea, R., & McKinney de Royston, M. (2020). *Handbook of the cultural foundations of learning* (p. 480). NY: Routledge.

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 Nasir, N., Lee, C., Pea, R., & McKinney de Royston, M. (Eds.), *Handbook of the cultural foundations of learning* (pp. 195–211). NY: Routledge.

National Academies of Sciences, Engineering and Medicine (2018). *How People Learn 2*, National Academies Press, Washington, DC.

Nathan, M. J. (2021). *Foundations of embodied learning: A paradigm for education*. NY: Routledge.

Newell, A. (1994). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Nichols, S. N., & Berliner, D. C. (2007). *Collateral damage: The effects of high stakes testing on America's schools*. Cambridge, MA: Harvard Education Press.

Niemi, H., & Nevgi, A. (2014). Research studies and active learning promoting professional competences in Finnish teacher education. *Teaching and Teacher Education*, *43*, 131–142.

Oakes, J. (2005). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.

OECD (2019), PISA 2018 *Assessment and Analytical Framework*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/b25efab8-en

Organisation for Economic Co-operation and Development. (2018). PISA 2018: *Social disparities*. OECD Publishing.

Osher, D., Kidron, Y., Brackett, M., Dymnicki, A., Jones, S., & Weissberg, R. P. (2016). Advancing the science and practice of social and emotional learning: Looking back and moving forward. *Review of Research in Education*, *40*(1), 644–681.

Packer, M., & Cole, M. (2020). The institutional foundations of human evolution, ontogenesis, and learning. In N. Nasir, C. Lee, R. Pea, & M. McKinney de Royston (Eds.), *Handbook of the cultural foundations of learning* (pp. 3–23). NY: Routledge.

Pea, R. D. (1994). Seeing what we build together: Distributed multimedia learning environments for transformative communications. *The Journal of the Learning Sciences, 3*(3), 285–299. https://doi.org/10.1207/s15327809jls0303_4

Pinkard, N., Erete, S., Martin, C. K., & McKinney de Royston, M. (2017). Digital youth divas: Exploring narrative-driven curriculum to spark middle school girls' interest in computational activities. *Journal of the Learning Sciences*, *26*(3), 477–516.

Rogoff, B. (2023). Mutually constituting, fractal: individual and cultural aspects of holistic process. *Review of Research in Education*, 47(1), 60–83.

Rosa, J. D. (2016). Standardization, racialization, languagelessness: Raciolinguistic ideologies across communicative contexts. *Journal of Linguistic Anthropology*, *26*(2), 162–183.

Rosa, M., D'Ambrosio, U., Orey, D. C., Shirley, L., Alangui, W. V., Palhares, P., & Gavarrete, M. E. (2016). *Current and future perspectives of ethnomathematics as a program*. Springer Nature.

Rueda, R. (2013). 21st-century skills: Cultural, linguistic, and motivational perspectives. In D. Alverman & N. Unrau (Eds.), *Theoretical models and processes of reading*, 6th Ed. (pp. 1241–1267). Newark, NJ: International Reading Association.

Saracho, O.N. (Ed.), *Handbook of Research Methods in Early Childhood Education.* Charlotte, N.C.: Information Age.

Schwartz, R., & Plass, J.L (2020). Types of engagement in learning with games. In J.L. Plass, R.E. Mayer, & B.D. Holmes (Eds.), *Handbook of game-based learning* (pp. 53–80). Cambridge, MA: MIT Press.

Scriven, M. (1967). The methodology of evaluation. In R.W. Tyler, R.M. Gagne, M. Scriven (Eds.), *Perspectives of curriculum evaluation,* (pp. 39–83). Chicago, IL: Rand McNally

Seitzer, H., Niemann, D., & Martens, K. (2021). Placing PISA in perspective: the OECD's multi-centric view on education. *Globalisation, Societies and Education*, *19*(2), 198–212.

Seymour, V. (2016). The human–nature relationship and its impact on health: A critical review. *Frontiers in public health*, *4*, 260.

Shapiro, L., & Spaulding, S. (2024). *The Routledge Handbook of Embodied Cognition* (2nd ed.). Routledge. https://doi.org/10.4324/9781003322511

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher, 15*(2), 4–14.

Shute, V., Rahimi, S., Smith, G., Ke, F., Almond, R. G., Dai, C. P., & Sun, C. (2021). Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games. *Journal of Computer Assisted Learning*, 37(1), 127–141.

Solano-Flores, G. (2019). Examining cultural responsiveness in large-scale assessment: The Matrix of evidence for validity argumentation. *Frontiers in Education, 4*(43).

Steele, D. M., & Cohn-Vargas, B. (2013). *Identity safe classrooms, grades K-5: Places to belong and learn*. Corwin Press.

Stevens, R. (2020). Locating children's interest and concerns: An interaction focused approach. In Nasir, N., Lee, C., Pea, R., & McKinney de Royston, M. (Eds.), *Handbook of the cultural foundations of learning* (pp. 212–229). NY: Routledge.

Stigler, J. W., & Hiebert, J. (2009). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. Simon and Schuster.

Taylor, E. V. (2009). The purchasing practice of low-income students: The relationship to mathematical development. *The Journal of the Learning Sciences*, *18*(3), 370–415.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT press.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and brain sciences*, 16(3), 495–511.

Turner, J. H. (2020). *On human nature: The biology and sociology of what made us human*. Routledge.

U. S. Department of Education, Office of Educational Technology (2023, May). *Artificial Intelligence and Future of Teaching and Learning: Insights and Recommendations*. Washington, DC.

Vakil, S. (2024). *Towards New Horizons of AI, Learning, and Equity in Education*. Spencer Foundation research report.

Valdés, G. (2004). *Learning and not learning English: Latino students in American schools*. NY: Teachers College Press.

Valencia, R. R. (2010). *Dismantling contemporary deficit thinking: Educational thought and practice*. NY: Routledge.

Volante, L., Klinger, D. A., & DeLuca, C. (2024). The rise and stall of standards-based reform. *Phi Delta Kappan*, *106*(2), 42–46.

Wang, T., & Cheng, E. C. (2022). Towards a tripartite research agenda: A scoping review of artificial intelligence in education research. In *Artificial Intelligence in Education: Emerging Technologies, Models and Applications* (pp. 3–24). Singapore: Springer.

Yaden, D. B., Rueda, R., Martinez, C., Mirzaei, A., Scott-Weich, B., Tardibuono, J., & Tsai, T. (in press). *A cross-linguistic perspective on studies using psycho-and micro-genetic design elements and methodology to assess young children's understandings of reading, writing and spelling.*