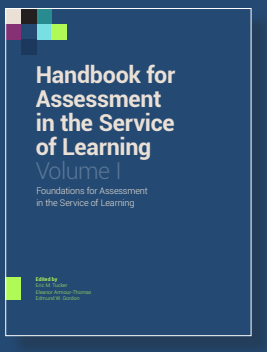
Designing for the Future: Toward an R&D Agenda to Promote Inclusive, Human-Centered Assessment Systems

Temple S. Lovelace, Orrin T. Murray, and Laura S. Hamilton

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:
Lovelace, T. S., Murray, O. T., & Hamilton, L. S. (2025). Designing for the future: Toward an R&D agenda to promote inclusive, human-centered assessment systems. 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.

Designing for the Future: Toward an R&D Agenda to Promote Inclusive, Human-Centered Assessment Systems

Temple S. Lovelace, Orrin T. Murray, and Laura S. Hamilton

Abstract

This chapter examines how inclusive human-centered design (IHCD) can transform educational assessment systems to better serve diverse learners. As the demands of education have expanded over time, assessment paradigms must likewise evolve beyond narrow academic measures to support holistic student development. The authors argue that assessment systems should be intentionally designed through IHCD principles that center the experiences and needs of learners, educators, and communities. The chapter explores three critical layers of an inclusive assessment R&D infrastructure: content that reflects modern understanding of development across domains, technological capabilities that enhance assessment utility, and broader policy support that ensures coherence and equity. Drawing on the Gordon Commission's Principles for Assessment in Service of Learning, the authors provide guidance for creating assessment systems that are transparent, authentic, culturally responsive, and actionable. The chapter addresses the potential benefits and risks of emerging technologies, particularly AI, while emphasizing the importance of stakeholder engagement in assessment design. By reimagining assessment through IHCD, education systems can move beyond current incoherent approaches toward paradigms that genuinely support learning, equity, and the multifaceted purposes of public education in preparing young people for education, careers, and civic life.

Introduction

The knowledge that an adolescent might have needed to successfully navigate young adulthood in the United States 150 years ago pales in comparison to what an adolescent must know and have faculty with today. And while the broad outlines might look familiar, what a school must do to prepare its students to succeed has significantly expanded in the last century and a half (Roschelle, Pea, Hoadley, Gordin & Means, 2000). Although the early struggles over the purpose of public schools have shifted away from the battlegrounds surrounding religion and morals, other social, cultural, and political concerns, including topics of race, gender, and social and emotional development, continue to roil the system of public education in the United States (Lampen, 2022; Mehta, 2025; Schwartz, 2021).

Perhaps the most consequential contemporary shift in how people talked about the purpose of education began with the publication of *A Nation at Risk,* which signaled a new front in the battle over the purpose of public schools (National Commission on Excellence in Education, 1983). As the United States' economic footing shifted from a manual labor focused workforce—rooted in manufacturing and agriculture—toward one that relied more on cognitive competencies (with an emphasis on technical skills in math and science) the last 50 years have seen increased emphasis on skills that directly supports this newer economic world standing (Hanushek & Woessmann, 2012; Zaber et al., 2019). For several decades, the idea that public schools serve the explicit purpose of developing vibrant citizens—people who vote in every eligible election, serve on juries, run for public office and are active at all levels in their communities—was at best implied, and, at worst largely forgotten by those charged with managing our education systems. Recent years have witnessed a renewed emphasis on cultivating citizens and promoting holistic development (Lee, White, & Dong, 2023; Vinnakota, 2019), but the purpose of public education remains contested, and current assessment and accountability policy does not, for the most part, reflect this broader vision (Hamilton & Martinez, 2024).

In keeping with the views and evidence presented throughout this chapter and volume, we assert that publicly funded schools must prepare young people to succeed and thrive in education settings (i.e., PreK–16+), careers, and civic and community life—though we acknowledge that schools cannot accomplish this on their own. These are ambitious objectives for education, and how we monitor and seek to address opportunities to reach these objectives—with the explicit support of assessment paradigms—must be grounded in the multifaceted purposes of

public education. Moreover, we argue that these assessment paradigms—many which do not currently exist—must rest on inclusive and human-centered designs. In offering assessment paradigms, as a bridging agent between our objectives and practices, it is critical to highlight that assessments (as with all tools or technologies) have both been designed to harm and have been used to harm in ways that were intentional and unintentional. The approach to assessments we offer here intentionally focuses this powerful tool explicitly in the service of ensuring that the benefits of assessment outweigh potential harms for all learners.

HCD and Inclusion

In some respects, the idea that the design of assessment paradigms has not been considered from the perspective of learners and educators making use of it seems nonsensical. Put another way, the concept of human-centered design (HCD), from the perspective of the words offered, appears redundant and obvious, for we have often overlooked the true meaning of HCD even as we have aimed to apply its principles. The status quo within design, prior to HCD-based movements, was elitist, and expert driven—where expertise centered on a canon cultivated by a small coterie of gatekeepers. Even as HCD gains popularity, it would be shortsighted to suggest the struggle in the design world is over; it is not. The rapid and wide adoption of HCD certainly opened the door for designers to consider the role that a more empathetic approach to inclusion, centering people first (e.g., the inclusion of Empathize in Stanford d.school's design thinking process or the Gather Information step in IDEO's process), and it has shifted the field in ways that the prior generations of designers might not recognize. For example, as the bulk of design work shifted from material objects to intangible objects like websites and applications for digital platforms, from a design perspective, a focus on beauty has shifted somewhat to focus on utility. As we consider the role of assessment paradigms, this lesson from inclusive, human-centered design can be instructive; for we must contend with what "utility" means in educational assessment and how that can impact the design, implementation, and use of the artifacts from those paradigms. For, in an inclusive and human-centered approach to assessment R&D, considering who is involved at every step is important to fully addressing the role of utility within assessment paradigms. The emphasis on utility signals that an inclusive, human-centered approach forces designers to consider the "for whom" their designs are intended for—who can, who will, and in what ways do the people who imbue them enter into the design conversation. Therefore, IHCD became an

opportunity for us to consider the question of whom does assessment design serve (e.g., "Why are we doing this? Why is it being done this way?").

As we look to HCD for developing assessments for education, interest in education addressing social justice, social and emotional development, equity and inclusion, though in retreat at the federal level in early 2025, still seems ripe for considering the questions, "for *whom* is this assessment designed?' as well as "for *what* is this assessment designed?"

The Evolution of HCD Ideologies

In *The Design of Everyday Things,* Norman describes HCD as "an approach that puts human needs, capabilities, and behavior first, then designs to accommodate those needs, capabilities, and ways of behaving." (Norman, 2013). Norman goes on to note that HCD starts with "a good understanding of people and the needs [of these people] that the design is intended to meet." What people need, implies function and use in ways that could outweigh a pure aesthetic, or even functional beauty. On its face, however, much like the Declaration of Independence or the U.S. Constitution, details about who the people whose needs, capabilities and behaviors are taken into account, matter.

The rise of what we now label human-centered design can be traced to a set of practices that helped to fuel the successes of David Kelley and his design firm, IDEO. While browsers like Netscape are seen as the principal point of leverage for popularizing the internet (McCullough 2018), the rapid expansion of user interface and user experience skills, which draw heavily on HCD, provide infinite use cases for why HCD matters to the wide use and thereby success of the internet. It is however, only recently that education researchers seriously sought to consider the concept of design.

The persistent failure of public schools to close opportunity gaps is but one telling marker for why we need to clarify that the kind of HCD for designing assessments we have in mind is explicitly inclusive. Where commercially focused designers are often rewarded to segment and focus on particular groups to the exclusion of others, such a design principle is incompatible (even if, from a practical perspective, such principles are ignored or pushed to the periphery) with compulsory public education—it is harder to pointedly exclude some students in how we think about teaching and learning, and thus assessments.

Building upon the solid foundation provided by researchers promoting the science of learning (Goldman & Lee, 2024; Osher, Cantor, Berg, Steyer & Rose, 2020), we believe that the assessment systems that are enacted in the future will need to be more responsive, relative to our current systems, to the assets and experiences students bring into learning environments. As such, these future assessment systems could be designed to offer educators timely and relevant data to inform a holistic, responsive approach to instruction. Even more, these future assessment systems should consider how caregivers and learners might also benefit from access to these data, especially as what we value as learning further extends to contexts outside of a brick and mortar school building. There are also lessons on the transformation needed from other fields. For example, the functional approaches, such as those found in medicine are another important consideration when thinking about IHCD (Cantor, 2021). A surgeon using rigid systems to assess patient risks, built on population averages will, on average, make care decisions that lead to positive patient outcomes, but treatment decisions on either side of the average will more often lead to fatal or catastrophic outcomes. And while a single instructional decision does not typically have the same effect as that of a surgeon, the accumulation of such decisions using assessment systems that are insensitive to some students mean that those students will more often than not receive inadequate instruction and be denied opportunities for professionally and personally rewarding pursuits.

In this chapter, we build on the work described in the earlier chapters to consider how to apply an inclusive human-centered approach to create and implement *systems* of assessment that center the needs of learners. We start by reviewing and elaborating upon the description of inclusive, human-centered design (IHCD) presented earlier in this volume. We then describe the key features of assessment systems, discuss how the IHCD principles can be enacted to develop coherent, learner-centered assessment systems, and briefly discuss the implications of advances in technology. We conclude by sharing some examples of promising developments in R&D that we believe will contribute to a future characterized by more inclusive, learner-focused assessment systems.

Applying ICHD to Systems of Assessment

Inclusive human-centered design (IHCD) is an approach to the design and development process that ensures that the process of design starts with an empathetic understanding of students' lived experiences and that the outcomes of design support a set of solutions that reflect their lived experiences. In a similar vein, assessment, as described by Mislevy (2019), is "situated in social contexts, shaped by purposes, and centered on students' developing capabilities for valued activities (p. 164)." The questions asked of inclusive approaches to human-centered design must also be asked of the development approaches used for assessment. If assessments are to be responsive, relevant, and provide the opportunity for reflection, then the methods and processes we use must follow the tenets laid forth for IHCD. These tenets, loosely summarized here and applied to assessment development, are that IHCD is about the inclusion of learners and educators into the process, "with whom," that it not only centers the purpose of the assessment process on the people most impacted, "for whom;" but that the context or environment in which the assessment process will be utilized matters, "for what". This process of centering these questions at the nexus of the assessment process is complex, yet it is a critical element to ensure we understand how assessment as a whole needs to evolve to better meet the needs of learners now, but certainly the learners in the future.

Before discussing the applications of specific IHCD principles to assessment and assessment systems, it is important to clarify how we are using these terms in the context of educational settings. "Assessment" can be defined as a "process of reasoning from evidence" (Foster & Piacentini, 2023, p.16) or, in a slightly more detailed description, "...the process and outcome of collecting and analyzing data to inform an interpretation, judgment, or decision about an attribute or property of an object, event, or phenomenon" (Ackerman et al., 2024, p.2). This definition applies to a variety of tools and approaches including large-scale, standardized achievement tests, curriculum-embedded assessments, and the informal data teachers collect almost constantly through their interactions with students. Additionally, by describing assessment as both a process and an outcome, the definition highlights the importance of design, development, administration, scoring, reporting, and the decisions that users make in response to results.

Regardless of the approaches used at each stage, assessment can serve various purposes including, but not limited to:

- Informing classroom instruction by providing information to educators and learners about learners' progress, assets and needs.
- Guiding learners' decisions about academic or career pathways.
- Selecting students into programs or institutions.
- Monitoring learning progress on a large-scale.
- Serving as a source of data for research that advances the empirical evidence base.
- Signaling to educators, caregivers, or other groups what kinds of learning outcomes they are expected to value and prioritize.

Of course, no single instrument can serve all these purposes effectively, and it is unwise to use a measure for a purpose other than the one (or ones) for which sufficient validity evidence has been gathered (AERA, APA, & NCME, 2014). Consequently, most young people will encounter numerous assessments, designed for a variety of purposes, each year as they make their way through the K–12 system and move into college, careers, or other pursuits.

Although some K–12 assessments are designed and validated for specific, relatively narrow purposes that do not typically influence students' broader educational experiences (for example, clinical tests to diagnose specific educational or mental health needs), many others influence what happens in classrooms and schools, whether or not they are designed for that purpose. These include assessments teachers create, curate, or modify to inform their instructional decisions; interim or benchmark assessments that districts might purchase to generate data regarding students' progress toward meeting state standards, and statewide accountability tests that are externally mandated but that often exert powerful effects on instruction (Faxon-Mills et al., 2013). Different assessments might send conflicting signals about what outcomes educators are expected to promote, and they might generate confusing or misleading data on student learning. As a result, their use can hinder educators' opportunities to provide a coherent instructional experience for students.

The concept of a *system* of assessment provides a way of thinking about how a set of assessments, designed for different purposes, might be integrated into a broad suite of tools and practices that sends aligned rather than conflicting signals to educators, students, and caregivers, and that is designed explicitly to support teaching and learning. Although there is no consensus definition of an "assessment system," nor a single clear framework to guide its development, a body of work conducted over the past two decades has converged around a vision put forth in the seminal 2001 volume, *Knowing What Students Know* (National Research Council, 2001). For this chapter, we draw on a definition offered in a recently published National Academy of Education volume called Reimaging Balanced Assessment Systems (Marion, Pellegrino, & Berman, 2024), which draws on *Knowing What Students Know* and subsequent scholarly work. The volume's editors and authors conceptualize balanced assessment systems as "intentionally designed to provide feedback to students and information for teachers to support ambitious and equitable instructional and learning opportunities" (Marion et al., 2024, p.2), with supports for high-quality formative assessment practices in classrooms along with large-scale district or state assessments that are designed to produce aggregate data for the purposes of monitoring, informing decisions about resource allocation or policy, and signaling expectations regarding instruction that will advance quality and equity. Coherence is a critical feature of this vision of balanced assessment systems, along with learner-centeredness, which we define as a relentless prioritization of improving student learning throughout all parts of the system (National Research Council, 2001; Marion et al., 2024).

Broadening our understanding of IHCD–the role of a coherent Assessment R&D Infrastructure

Efforts to make assessment paradigms more inclusive and human-centered can not start once the assessment arrives in front of an educator or learner. Given the role of IHCD in ensuring that value, effectiveness, and use are core components of assessment development, expanding its application to the full research and development (R&D) process is a critical step. Traditionally, assessment developers have engaged in R&D processes that reflected a relatively narrow range of expertise and that did not typically emphasize the role that assessments play in broader systems. In order to realize an assessment system that is inclusive and human-centered, it must be constructed on a multi-faceted infrastructure. The

infrastructure for assessment R&D is one that must consider not only the creation of innovative tools, but also the sustained use of those tools and the local and federal policy needed to shape and support it. This inclusive and human-centered infrastructure has to honor the experiences of the learners, educators, caregivers, school leaders, and others, but it must also prioritize an orientation towards the future of the learning and measurement sciences.

We offer that there are three layers within an inclusive, human-centered R&D infrastructure that are worthy of note. The first, focuses on the *substance or content* of our assessment systems, such as how they reflect a modernized view of child development across multiple domains, the degree to which they support culturally relevant experiences that promote the individualization of education, and their responsiveness to the growing evidence base stemming from advances in the learning sciences. The second layer contains the *technological capabilities* that expand the utility of those innovations, such as leveraging digital innovations (from capturing still and moving images in learning environments to using AI to support recommendations for pedagogical development or to tailor student development) and the use of multi-modal components within assessment. The last layer reflects the *broader landscape* of an IHCD-aligned assessment R&D infrastructure, which includes the local context and state and federal policy support needed to ensure a robust R&D experience, and how we creatively resource this expanded infrastructure. The position we are staking out moves past the incoherence of the current system of assessments by insisting that, going forward, assessments are anchored in an inclusive and human-centered R&D paradigm. Therefore, it will be critical that developers of assessments frame their ICHD with learner needs driving the alignment of educators, administrators and policy makers towards a more coherent and learner-focused structure.

Exploring the Role of an IHCD-Informed Infrastructure on the Learning and Measurement Sciences

The current PreK-16+ education system has long been impacted by the policies set forth for assessment and for standards-aligned learning. For decades, scholars have called for investigating a more comprehensive approach to learning that considers not only what academic content a learner should know, but also the skills that attend to their social and emotional competencies (Lovelace, McMurtry, and Kendall-Brooks, 2024). Even more, there is the consideration for how learners

may apply their skills for civic engagement, such as considering their role as change agents, and expanding the application of their skills to advocacy-related fields (Jagers et al., 2017). These considerations, which align to a call for new approaches in the learning sciences that consider the role of identity, agency, and wellbeing provide a compelling argument for an inclusive, human-centered assessment system that can support the creation and sustaining of learning environments that provide the opportunity for growth in these areas. However, to do so convincingly, the underlying assessment R&D infrastructure must be reflexive. For example, our current assessment R&D infrastructure does not equitably support the assessment of essential skills (i.e., 21st-century skills) at the same rate as skills associated with standards-based reading, math, and science -based content. Even more, there is less cohesiveness in the area of social and emotional competency assessment, with over 200 frameworks that are currently focused in this space and an insufficient evidence base to inform appropriate use of such assessments (Assessment Work Group, 2019; Berg et al., 2017; Harvard University, undated; Casillas, Roberts, & Jones, 2022). A related challenge is that most currently available assessment tools are designed to measure student learning in isolated ways, such as through separate math, reading, and social and emotional competency measures. As we consider how an IHCD infrastructure could better reflect what we know about how people learn, it will be critical to engage in R&D to advance assessment content and methods in ways that are more comprehensive and that reflect a more holistic understanding of student experience.

As we consider how we might prioritize these efforts going forward, the field must consider the relationship that assessment paradigms have to our curriculum and instruction infrastructure. The Gordon Commission's Principles suggest that this relationship is bi-directional and reflexive; one informs the other in a cyclical and continuous manner (Gordon & Pellegrino, 2013). Thus, the infrastructure to support assessment R&D must grow to support that. For example, our processes for measurement, our assessment practices and culture, as well as content, must be continually reviewed and updated to support new learning paradigms.

How Assessment Systems can Benefit from Application of the Principles for Assessment in Service of Learning

The Principles for Assessment in Service of Learning, a product of the Gordon Commission, offer a promising framework for informing both the design and implementation of assessment systems (Gordon & Pellegrino, 2013). In this section, we offer guidance drawn from these principles that we believe will help maximize the likelihood that systems of assessments are coherent and learner-centered and that they reflect the values and priorities of the groups who will be most affected by them. This guidance is not exhaustive but instead intended to highlight a small number of ideas that can inform current assessment system design efforts and future R&D, which we discuss later in this chapter.

Engage educators, learners, and caregivers in assessment system design. Principle 1 emphasizes the importance of transparency, which is relevant not just to assessments themselves but to the broader systems in which assessments are embedded. An inclusive approach to informing assessment system design should provide members of affected groups—mainly educators and learners, but also administrators, caregivers, and other community members—opportunities to inform all aspects of the system. This is not to suggest that professional assessment expertise is not needed or that members of these groups would make all decisions. Rather, application of Principle 1 would involve engaging with groups in appropriate ways, such as by inviting and incorporating input on initial design decisions and providing multiple opportunities for review and feedback on interim products. Engagement opportunities should be customized to the needs and assets of each group and should incorporate specific strategies and mechanisms for reviewing and addressing feedback (Hamilton & Landl, 2023). Frameworks associated with design-based research and user-focused R&D, which emphasize iterative cycles of testing, feedback, and adaptation, can be especially valuable (Armstrong, Dopp, & Welch, 2020; Design Based Research Collective, 2003; Hamilton & Murray, 2023; Schneider, 2023). Importantly, such engagement must go beyond giving educators or other groups "voice" and instead must allow these groups to exert agency and influence in ways that are meaningful and responsible.

Incorporate authentic activities into assessment design. One of the reasons efforts to adopt balanced assessment systems have often fallen short of their goals is that they have been overly influenced by externally mandated, high-stakes testing. The Elementary and Secondary Education Act (ESEA) has relied on statewide assessments as a key policy lever, resulting in tests that have led to a narrowing of curriculum and instruction, an overly constrained notion of what constitutes "effective" instruction, and a tendency for educators to rely heavily on commercially available interim or benchmark testing systems that aren't always designed to support high-quality teaching (Evans, 2024). Systems that address principle 2, which refers to progress, outcomes and processes that are transferable, must prioritize assessment tools and practices that do not always lend themselves to large-scale administration. These might include more sophisticated items and tasks such as scenario-based or portfolio assessments, along with informal assessment approaches that skilled teachers use to gauge their students' understanding and needs through their interactions in learning environments.

This is not to suggest that large-scale assessments cannot play a valuable role in systems of assessment. The National Assessment of Educational Progress (NAEP), for instance, provides a mechanism for monitoring performance and progress on a large scale and for exploring relationships between student achievement and other factors such as economic or instructional resources. Large-scale assessment data can be especially powerful when connected to information about students' learning opportunities and access to resources (National Academies of Sciences, Engineering, and Medicine, 2019). A set of assessments for monitoring progress at the state or district level could play an important role in an assessment system but must be designed and implemented in ways that prevent such tests from having significant influence over instruction, such as by refraining from calculating individual student scores or imposing consequences based on performance. In this way, large-scale tests can serve complementary purposes to other assessments in the system.

Anchor systems in whole-child conceptualizations of learning. A large and growing body of evidence on how people learn has advanced the assessment field's understanding of the inseparability of academic, social, and emotional development and the ways in which learning is influenced by one's cultural context and assets (Goldman & Lee, 2024). Educators have largely endorsed the goals of whole-child education (Hamilton et al., 2024; Rikoon et al., 2024). This

science-based perspective can help assessment systems align to principle 3, which acknowledges the importance of enabling learners to draw on their own knowledge and experiences while participating in assessment, and principle 4, which calls for assessment that supports multiple aspects of learners' development. Advances in classroom assessment tools and practices offer opportunities to incorporate a whole-child perspective. For instance, some digital assessment tools provide data not just on academic learning but on engagement and other related constructs (Linzarini & da Silva, 2024). More holistic approaches to measuring student progress in large-scale assessment systems are also being explored by states, districts, and advocacy groups (Domaleski et al., 2023; Hamilton & Martinez, 2024; National Urban League & UNIDOS, 2022). Of course, adopting a whole-child perspective carries some risks and must be undertaken thoughtfully. For instance, incorporation of social and emotional competency measurement could lead to inappropriate uses of such data and to a reliance on measures and construct definitions that are not culturally responsive to all groups (Assessment Work Group, 2019; Jagers et al., 2018). An IHCD approach provides one bulwark against misguided efforts by helping to ensure broad input on what aspects of learners' development should be measured, how they should be measured, and how the resulting data should be used.

Recognize how assessment systems can communicate and model expectations. The theory of standards-based reform, which informed the design of state accountability systems over the past few decades, emphasized the idea that clearly specified content standards could help educators understand what student learning outcomes they should promote but without offering prescriptions regarding pedagogy (Smith & O'Day, 1991). Research has shed light on ways in which this theory of action has failed to be realized and has found that it is often the tests rather than the standards that drive instruction (Koretz, 2008; Koretz & Hamilton, 2006). The fundamental problem with this view is that large-scale assessments are designed to accommodate broad content coverage in the context of limited administration time and funding, which in turn prioritizes item types that are poor models of instruction (Darling-Hammond & Adamson, 2010). An assessment system aligned with principle 5, which calls for assessments that can serve as models of learning, would include classroom assessment tools and practices that engage students in activities that reflect important, real-world activities.

Monitor and improve system quality based on the specific purposes of each component of the system. Principle 6 emphasizes the importance of clearly specifying the purpose of an assessment and ensuring that it supports appropriate inferences relevant to that purpose. Additionally, this principle echoes widely accepted assessment guidance regarding the need for multiple sources of evidence (AERA, APA, NCME, 2014). Systems of assessment typically incorporate several components, each of which is designed to serve a specific purpose (e.g., informing instruction, monitoring equity on a large-scale). Those who design and implement assessment system components need to engage in a process of continuous review and improvement to ensure that decisions based on each component are appropriate and warranted based on the quality of evidence those components produce and that the system as a whole generates credible, coherent data to inform improvement. Application of IHCD to this continuous improvement process involves seeking feedback from all relevant groups to understand the inferences and decisions they are making based on results and to elicit their recommendations for modifications.

Such an approach also helps advance principle 7, which emphasizes the need for assessment results to be tied to actionable feedback that informs decisions. Systems of assessment are typically intended to inform a variety of decisions, and it is critical that users of information from such systems understand how to make sense of information from different components of the system and how to use this information to make decisions that improve teaching and learning. This means being clear on what kinds of decisions are typically not warranted, such as the use of large-scale, standardized test scores to inform instruction (Marion, Pellegrino, & Berman, 2024).

The Role of Emerging Technologies in an IHCD-Informed R&D Infrastructure

Several scholarly, societal, and market-based trends are influencing the work of measurement professionals and the design and implementation of assessment tools and practices. Returning to Mislevy's (2019) conceptualization of assessment as "situated in social contexts, shaped by purposes, and centered on students' developing capabilities for valued activities (p. 164)," it is clear that the growing recognition of assessment's role in propagating systemic inequities must influence how we work to situate assessment in learners' diverse, multifaceted contexts. Additionally, the purposes of assessment are changing, and society's

notions regarding what activities are "valued" will inevitably evolve in response to a changing economic and civic landscape. Other chapters in this volume discuss some of these factors in detail. In this section, we focus on one potentially influential set of conditions—namely, advances in applications of technology to assessment and, in particular, the growing availability and influence of generative AI. The role of technology in an IHCD—infrastructure is likely to become more significant over time, in that most advances in assessment systems are often leveraging a more technology-enabled approach to how we assess student learning and educational outcomes. The role of technological capabilities as a way to improve our assessment paradigms will always and inevitably influence the assessment field, yet how we approach their arrival and early use is an area that will require significant improvement and a much more solid evidence base than currently exists.

An R&D infrastructure that supports IHCD-aligned practices includes a place where curiosity and early adoption can occur in a safe, cautious way. These technologies will influence our assessment paradigms, even if those actors are "external" to the organizations in which assessment R&D is their primary focus. As seen in the current market for curriculum and instructional tools, venture capital, private equity, and those in non-educational fields have exerted increasingly significant influence on the learning technologies used in schools. One question we may want to answer is how might the assessment field expand to include a more robust place within its infrastructure to hold this curiosity and faster-paced exploration in service of leading edge assessment R&D?

Responsibly Leveraging Advances in AI

The release of ChatGPT in November 2022 launched a wide-ranging debate that has raged and morphed ever since among advocates, skeptics, and everyone in between. A narrower version of this debate has been prominent within the measurement community, consuming pages of journals and books along with panel discussions at conferences. These conversations are valuable, particularly when they include representatives from diverse groups and with varying experiences and viewpoints. Given how much we don't know, and how rapidly the generative AI landscape is advancing, it is critical for assessment system designers to avoid a knee-jerk rush to "innovation" or a panicked avoidance of all things AI. Instead, an approach that combines a cautious embrace of new ideas

with an iterative, user-focused approach to testing and refining tools is likely to offer the best chance to create systems that promote the desired outcomes while limiting harm. School district leaders and other experts have been issuing guidance regarding appropriate AI use[1], though most of that guidance devotes relatively little attention to assessment. In the rest of this section, we do not attempt to lay out the many ways in which generative AI might improve or detract from high-quality assessments. Instead, we emphasize the ways in which IHCD can support responsible applications of AI to assessment systems.

Continue to Build the Evidence Base

Despite the vigorous enthusiasm that generative AI has prompted in many circles, evidence to inform decisions about how it should be incorporated into assessments is, with a few exceptions (e.g., methods for automated essay scoring), quite nascent. Because the adoption of AI for assessment is inevitable, an inclusive assessment R&D infrastructure must prioritize work to test the effects of promising AI applications on score validity, reliability, and fairness. Innovation in assessment is clearly needed, but iInnovation for the sake of innovation is unlikely to yield the intended benefits and could result in significant harms.

Ensure Stakeholder Understanding of, and Support for, AI Applications

An IHCD approach is well-suited to addressing the potential benefits and risks of AI. By engaging all relevant groups in conversations designed to enhance their understanding of AI while also eliciting their concerns and hopes regarding AI applications, developers can help ensure responsible adoption. It is especially crucial for these groups to understand the likely implications of incorporating (or not incorporating) AI for a particular assessment feature. Additionally, the increasing availability of tools to support development of "home-grown" AI tools[2], increases the urgency of engaging in collaborative conversations and reaching a collective agreement on specific AI applications.

1 For a particularly promising example, see this guidance developed by Santa Ana Unified School District: https://www.sausd.us/cms/lib/CA01000471/Centricity/Domain/6/DRAFT%20-%20SAUSDs%20AI%20Compass%20Your%20Guide%20to%20Navigating%20an%20AI%20World%209.pdf

2 https://openai.com/index/introducing-gpts/

Explore How AI Might Promote Coherence

One promising approach to pursuing the vision of coherent assessment systems described earlier is through steps that diminish the likelihood that large-scale assessments will negatively affect teaching and learning (Marion et al., 2024; Hamilton & Martinez, 2024). Applications of AI to large-scale assessments could contribute to this goal in a few ways, including through improved approaches to adaptivity that could reduce test length and through new techniques for personalizing assessments that could potentially improve alignment with other parts of the assessment system.

Consider Benefits and Drawbacks of Using AI to Promote Equity and Cultural Responsiveness

As we consider the role of the seven principles of assessment in the service of learning, advanced technical capacities have a role to play in advancing a more individualized and equitable learning environment. For example, in order to advance an assessment system that allows for individualization that is responsive to that learner's specific needs (Principles 3,5,6) or an assessment system that provides more robust and relevant results (Principle 7), the infrastructure in which it sits must be more interoperable; must be more secure; and, must allow for more advanced analytics that are combined with greater attention to data literacy support for its users. However, in order to ensure that these capabilities don't duplicate the harm that has been evidenced in the past, as discussed earlier in this chapter, this infrastructure must consider how it simultaneously advances a more inclusive and human-centered approach.

IHCD Applications to Policy, Practice, and Use: The Role of Establishing an Assessment Culture Rooted in Transformation

Systems of assessment, as previously mentioned in this chapter, are not only shaped by advances in the sciences, or the application of new technologies, they are also molded by the cultural and political environments in which they are situated. In addition, the infrastructure needed to support these advances will be more robust if it is guided and resourced at the state and federal levels. This intentional shaping of assessment paradigms, even through policy, can help to ensure that there is widespread use and application of these principles at a larger scale. While there is certainly a directionality when it comes to how policy shapes assessment practices, including the factors that support effective

implementation, by elevating a more inclusive and human-centered approach into how we craft policy, it is possible to occasion a more equitable and effective paradigm for assessment.

There are several considerations that can be advanced when thinking about the role of ICHD in how we shape the culture of assessment, and we offer a few levers for consideration. For example, how we support educator and learner use, the expansion of our ideologies within our generation and analysis of assessment data, and how we might advocate for a more inclusive approach in the design and application of policy can provide a compelling set of use cases for how we may move forward in ensuring that ICHD-aligned assessment paradigms are a fixture in the future of education.

We anticipate that the marketplace of integrated digital curriculum and assessment tools will continue its rapid growth and that educators will face an increasingly complex set of decisions regarding what tools to adopt, in what contexts, and for what purposes. Those who prepare and support educators—including classroom teachers as well as administrators who often make purchasing and adoption decisions—should develop research-based guidelines for adoption and use. Such guidelines should address questions such as how educators can make sense of the reams of data generated, how they can achieve policy objectives such as disaggregation of data and personalization of instruction to meet individual student needs, and how to interact effectively and appropriately with chatbots and search tools. Aggressive marketing of software and other tools for classroom and school use has created confusion and hindered adoption of balanced assessment (Hamilton & Martinez, 2024) and these challenges are only likely to grow.

Some Promising Efforts toward Inclusive Assessment System Design

In recent years, efforts to make sure that the R&D process, including assessment-related R&D, centers inclusion and accessibility have grown. These efforts span philanthropic organizations, such as the Chan Zuckerberg Initiative (CZI), The Spencer Foundation, and the Gates Foundation; non-profit research organizations, such as the American Institutes for Research (AIR) and the Advanced Education Research and Development Fund (AERDF); and even federal-level programs, such as National Center for Advanced Development in Education (NCADE) at the Institute for Education Sciences (IES) and the National Science Foundation (NSF). The Alliance for Learning Innovation convened a group of researchers, practitioners,

policymakers, and others to develop recommendations related to inclusive R&D; these include investing in collaborative work that honors the perspectives of diverse groups and using measurement strategies that generate more accurate and useful data for all participants (Alliance for Learning Innovation, 2024).

As a concrete example of ICHD, the Measures for Early Success Initiative, housed at the MDRC (Manpower Demonstration Research Corporation) seeks to reimagine the assessment landscape, especially the PreK assessment systems that operate at local and state levels. Through this process, they are leveraging a clear human-centered design approach, called participatory R&D, that ensures that families, educators, and learners themselves are a part of the assessment design process—even in tools that are being used at scale across the nation. This year-long process requires that development teams meet on a consistent basis with educators, caregivers, and learners, and that they weave the information gleaned from them directly into the tools they are creating.

Conclusion

As we close this chapter, we invite readers to consider what priorities might be important for the field in service of seeking an evolution to a more inclusive, human-centered set of assessment paradigms and the implications of taking on this new lens in support of a more equitable set of assessment paradigms for education. Achieving the ambitious vision described in this chapter and in the rest of the volume will require a collaborative approach that engages educators, policymakers, education support providers, measurement experts, families, and, of course, the young people who have the greatest stake in the outcomes of these efforts.

References

Ackerman, T.A., Bandalos, D.L., Briggs, D.C., Everson, H.T., Ho, A.D., Lottridge, S.M., Madison, M.J., Sinharay, S., Rodriguez, M. C., Russell, M., von Davier, A.A. and Wind, S.A. (2024). Foundational competencies in educational measurement. *Educational Measurement: Issues and Practice.* https://onlinelibrary.wiley.com/doi/10.1111/emip.12581

Alliance for Learning Innovation (2024). *Inclusive Education R&D.* https://fas.org/wp-content/uploads/2024/05/ALI-Taskforce-Brief-Inclusive-RD.pdf

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

Armstrong, M., Dopp, C., & Welsh, J. (2020). *Design-based research.* https://edtechbooks.org/studentguide/design-based_research

Assessment Work Group. (2019). *Student social and emotional competence assessment: The current state of the field and a vision for its future.* Chicago, IL: Collaborative for Academic, Social, and Emotional Learning. https://measuringsel.casel.org/wp-content/uploads/2019/09/AWG-State-of-the-Field-Report_2019_DIGITAL_Final.pdf

Berg, J., Osher, D., Same, M. R., Nolan, E., Benson, D., & Jacobs, N. (2017). I*dentifying, defining, and measuring social and emotional competencies.* American Institutes for Research. https://www.air.org/sites/default/files/2021-06/Identifying-Defining-and-Measuring-Social-and-Emotional-Competencies-December-2017-rev.pdf

Burrus, S. H. Rikoon, & M. W. Brenneman. (Eds.). A*ssessing Competencies for Social and Emotional Learning: Conceptualization, Development, and Applications.* New York: Routledge. https://doi.org/10.4324/9781003102243

Cantor, P. (2021). All children thriving: A new purpose for education. *American Educator, 45*(3), 14–26. https://files.eric.ed.gov/fulltext/EJ1322300.pdf

Casillas, A., Roberts, B., & Jones, S. (2022). An integrative perspective on SEL frameworks. In J. Burrus, S. H. Rikoon, & M. W. Brenneman (Eds.), *Assessing Competencies for Social and Emotional Learning: Conceptualization, Development, and Applications* (pp. 9–27). Routledge. https://doi.org/10.4324/9781003102243-3

Darling-Hammond, L.& Adamson, F. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st-century standards of learning.* Stanford Center for Opportunity Policy in Education.

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5–8. http://www.jstor.org/stable/3699927

Domaleski, C., D'Brot, J., Pinsonneault, L., Gong, B., & Brandt, C. (2023). *The path forward for school accountability: Practical ways to improve school accountability systems now. Center for Assessment.* https://www.nciea.org/wp-content/uploads/2023/06/Path-Forward-For-School-Accountability-FINAL.pdf

Evans, C. (2024). Stop trying to use commercial interim tests for instructional purposes: A case for "strategic abandonment." *CenterLine Blog, Center for Assessment.* https://www.nciea.org/blog/strategic-abandonment-and-interim-assessment/

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

Foster, N. and M. Piacentini (Eds.). (2023). *Innovating assessments to measure and support complex skills.* OECD Publishing. https://doi.org/10.1787/e5f3e341-en

Goldman, S., & Lee, C. (2024). Human learning and development: Theoretical perspectives to inform assessment systems. In Marion, S. F., Pellegrino, J. W., & Berman, A.I. (Eds.), *Implementation and use of balanced assessment systems* (pp.48–29). National Academy of Education.

Gordon, E. W., Pelligrino, J. (Eds.). (2013). Technical Report of the Gordon Commission on the Future of Assessment in Education. *Educational Testing Service.* https://www.ets.org/Media/Research/pdf/gordon_commission_technical_report.pdf

Harvard University (undated web page). *Explore SEL.* http://exploresel.gse.harvard.edu/

Hamilton, L.S., & Landl, E. (2023). Improving accountability through authentic community engagement. *CenterLine Blog, Center for Assessment* https://www.nciea.org/blog/improving-accountability-through-authentic-community-engagement/

Hamilton, L.S., & Martinez, J.F. (2024). Policy influences on ambitious classroom instruction, assessment, and learning. In Marion, S. F., Pellegrino, J. W., & Berman, A.I. (Eds.), *Implementation and use of balanced assessment systems* (pp.274–308). National Academy of Education.

Hamilton, L.S., & Murray, O. (2023). *Accelerating progress in U.S. education: Key lessons from other federal R&D investments in technology and innovation.* American Institutes for Research. https://www.air.org/resource/brief/accelerating-progress-us-education-key-lessons-other-federal-rd-investments

Hamilton, L.S., Olivera-Aguilar, M., & Rikoon, S.H. (2024). *Cultivating civic learning and engagement in U.S. schools. American Institutes for Research.* https://www.air.org/sites/default/files/2021-06/Identifying-Defining-and-Measuring-Social-and-Emotional-Competencies-December-2017-rev.pdf

Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth, 17*(4), 267–321. http://www.jstor.org/stable/23324889

Jagers, R., Lozada, F. T., Rivas-Drake, D., & Guillaume, C. (2017). Classroom and school predictors for civic engagement among Black and Latino middle school youth. *Child Development, 88*(4), 1125–1138. https://doi.org/10.1111/cdev.12871

Jagers, R.J., Rivas-Drake, D., & Borowski, T. (2018). *Equity & social and emotional learning: A cultural analysis.* Collaborative for Academic, Social, and Emotional Learning. https://drc.casel.org/uploads/sites/3/2019/02/Equity-Social-and-Emotional-Learning-A-Cultural-Analysis.pdf

Koretz, D. (2008). *Measuring up: What educational testing really tells us.* Harvard University Press.

Koretz, D., & Hamilton, L.S. (2006). Testing for accountability in K–12. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp.531–578). American Council on Education/Praeger.

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

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

Linzarini, A., & da Silva, D.C. (2024). *Innovative tools for the direct assessment of social and emotional skills.* Organisation for Economic Co-operation and Development. https://www.oecd.org/education/innovative-tools-for-the-direct-assessment-of-social-and-emotional-skills-eed9bb04-en.htm

Lovelace, T.S., McMurtry, T., & Kendall-Brooks, L. (2024). Equitable and culturally relevant social and emotional competency assessment. In J. A. Durlak, C. E. Domitrovich, R. P., & Mahoney (Eds.), *Handbook of Social and Emotional Learning: Research and Practice.* (2nd. ed.). New York: Guilford.

Marion, S. F., Pellegrino, J. W., & Berman, A.I. (Eds.), *Implementation and use of balanced assessment systems.* National Academy of Education.

McCullough, B. (2018). *How the Internet Happened: From Netscape to the iPhone.* Blue Cypress Books.

Mehta, J. (2025). Education Dept. warns schools: Eliminate DEI programs or lose funding. *NPR*. https://www.npr.org/2025/04/03/nx-s1-5350978/trump-administration-warns-schools-about-dei-programs

Mislevy, R. J. (2019). Advances in Measurement and Cognition. *The ANNALS of the American Academy of Political and Social Science, 683*(1), 164–182. https://doi.org/10.1177/0002716219843816

National Academies of Sciences, Engineering, and Medicine. (2019). *Monitoring educational equity.* National Academies Press. https://nap.nationalacademies.org/catalog/25389/monitoring-educational-equity

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* U.S. Government Printing Office.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment.* National Academy Press.

National Urban League & UNIDOS (2022). *Education assessment, accountability, & equity.* https://nul.org/sites/default/files/2023-04/FOAA_Final%20Phase%201%20Report_April2023_final%20.pdf

Norman, D. (2013). *The Design of Everyday Things*. Basic Books. https://doi.org/10.15358/9783800648108

Osher, D., Cantor, P., Berg, J., Steyer, L., & Rose, T. (2020). Drivers of human development: How relationships and context shape learning and development. *Applied Developmental Science, 24*(1), 6–36. https://doi.org/10.1080/10888691.2017.1398650

Rikoon, S.H., Hamilton, L.S., & Olivera-Aguilar, M. (2024). *Social and emotional learning in U.S. schools.* American Institutes for Research. https://www.air.org/sites/default/files/2024-04/social-emotional-learning-US-schools.pdf

Roschelle, J. M., Pea, R. D., Hoadley, C. M., Gordin, D. N., & Means, B. M. (2000). Changing how and what children learn in school with computer-based technologies. *The Future of Children,* 76–101. https://doi.org/10.2307/1602690

Schneider, M. (2023, February 22). Innovation in the education sciences (the new IES). *IES Blog*. https://ies.ed.gov/director/remarks/02-02-2023.asp

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

Smith, M., & O'Day, J. (1991). Systematic school reform. In S. Fuhrman and B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233–268). Falmer Press.

Vinnakota, R. (2019). *From civic education to a civic learning ecosystem: A landscape analysis and case for collaboration.* Red & Blue Works. https://rbw.civic-learning.org/wp-content/uploads/2019/12/CE_online.pdf

Zaber, M. A., Karoly, L. A., & Whipkey, K. (2019). *Reimagining the workforce development and employment system for the 21st-century and beyond.* RAND Corporation. https://doi.org/10.7249/RR2768