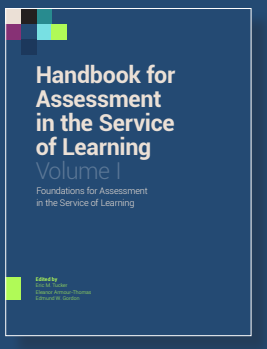
Measure What Matters for Teaching and Learning, and Measure It Well: Considering Design Principles and Approaches

Eric M. Tucker and Eleanor Armour-Thomas

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:
Tucker, E. M., & Armour-Thomas, E. (2025). Measure what matters for teaching and learning, and measure it well: Considering design principles and approaches. 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.

Measure What Matters for Teaching and Learning, and Measure It Well: Considering Design Principles and Approaches

Eric M. Tucker and Eleanor Armour-Thomas

Assessment must be reconceived as a tool for learning, not simply a tool for measurement (Baker & Gordon, 2014; Bennett, 2017; Shepard, 2019). For too long, tests have been used primarily to rank and sort students, reflecting aspects of what is—their current achievement—without guiding efforts to inform and improve what could be with regard to learning and development (Gordon, 1995, 2020, 2025; Gordon & Rajagopalan, 2016; Shepard, 2021). Volume I contributors argue that this paradigm must give way to approaches that illuminate how learning happens and how instruction can improve (Tucker & Armour-Thomas, 2025). If we hope to transform assessment systems, we must transform how we design, develop, and orient assessments–with clear purpose, strong grounding in how students learn, and a commitment to efficacy, quality, and usefulness in classrooms and learning contexts (Baker, Everson, Tucker, & Gordon, 2025; Huff, 2025; National Research Council, 2001; Pellegrino, 2025). The chapters in this section lay a foundation for human-centered and balanced assessment systems that foster deeper learning for every learner and address the real constraints that come with such a challenge.

Baker and colleagues (Baker et al., 2025) articulate design principles: be transparent about what is assessed, how, and to what ends; make purpose and focus explicit—clarifying intended outcomes, indicators of progress, and the learner processes that should transfer across contexts; and engineer tasks that support motivation, attention, and metacognition. They further urge designers to model trajectories over time; link assessment to feedback and adaptive instruction; accommodate learner variation with fair tasks and supports; and underwrite all claims with accessible evidence of quality and validity aligned to intended uses.

Pellegrino makes the case for assessment innovation, framing assessment as a process of reasoning from evidence and presents three arguments. First, we must "measure what matters, not just what is easy to measure" (Pellegrino, 2025). This means complex competencies—cognitive, socio-cognitive, and socio-emotional skills—that are essential for success in the future (Darling-Hammond, Herman, Pellegrino, Abedi, Aber, Baker, Bennett, Gordon, Haertel, Hakuta, Ho, Linn, Pearson, Popham, Resnick, Schoenfeld, Shavelson, Shepard, Shulman, & Steele, 2013; Foster & Piacentini, 2025). Second, we need design and delivery approaches that leverage modern technologies and tasks to capture richer evidence of learning (Huff, 2025; Linn, Baker, & Dunbar, 1991). Third, value proposition innovation in service of educators and learners is only as powerful as the evidence that underwrites it. We must apply rigorous standards of validation, reliability, and fairness to fulfill an obligation to add value to learners and educators (Chatterji, 2025; Pellegrino, 2025). In short, the promise of innovative assessment is inseparable from a robust evidentiary foundation; without trustworthy data, even the most imaginative assessment tasks are unlikely to serve learning.

Foster and Piacentini (2025) recognize that 21st-century education demands assessments of more complex skills—and that innovation is needed *across* the entire development process, from defining what to measure, to task design, scoring, and reporting (Darling-Hammond & Adamson, 2014; Marion, Pellegrino, & Berman, 2024). Piacentini and Foster propose five design principles for next-generation assessments, each grounded in research on how people learn. A student working through an extended task with built-in feedback is simultaneously being assessed *and* taught.

Lovelace, Murray, and Hamilton (2025) argue that transforming assessment requires rethinking the research and development (R&D) infrastructures. To counter system incoherence, anchor assessment in inclusive, human-centered design (Lovelace, Murray, & Hamilton, 2025). This means that assessment innovation should start with the needs and variation of learners at the center, honoring students' varied cultural and developmental contexts, and involve educators and communities in the design process. Lovelace and colleagues (2025) propose a three-layer agenda—modern, culturally relevant content; tools (e.g., simulations, AI-driven feedback); and enabling policy.

Huff (2025) observes that educational assessment stands at a "critical inflection point". The conventional approaches to test design that prevailed for decades, methods rooted in an era of norm-referenced tests used mainly to rank-order students, are no longer sufficient (Marion et al., 2024; Shepard, 2000). Today, there is growing demand for assessments that are "authentic, informative, actionable, engaging and accessible for all students" (Huff, 2025). First, the constructs we aim to measure have grown more complex–for example, mathematical practices, scientific inquiry skills, or collaborative problem-solving rich domains. Second, stakeholders expect assessments to serve multiple purposes (from guiding classroom teaching to informing accountability) without causing test overload, which means a single assessment design often must support several interpretations and uses. Third, there is heightened scrutiny of quality and fairness, pressuring developers to provide validity evidence and to design with rigor and transparency. Huff (2025) recommends Principled Assessment Design approaches, including evidence-centered design, to ensure task–claim coherence and explicit validity from the outset.

If we do not begin with design principles that privilege purpose, efficacy, coherence, balance, and evidence-based design, we risk innovating in name only–producing new solutions that replicate old problems. But as the chapters argue, if we design with integrity and intentionality, assessments have the potential to become powerful levers to improve and evaluate learning at all levels of the educational system—classroom, school, district, and state. The knowledge base on learning, cognition, and measurement has never been stronger, and the urgency for more effective, more humane assessment has never been clearer (Gordon, 2025; Mislevy, 2018; National Research Council, 2001; Penuel & Watkins, 2019). By centering design principles that prioritize learning, we–researchers, policymakers, educators, and assessment developers alike–can work together to build assessment systems that empower every student to learn and develop.

References

Baker, E. L., & Gordon, E. W. (2014). *From the assessment of education to the assessment for education: Policy and futures. Teachers College Record, 116,* 1–24.

Baker, E. L., Everson, H., Tucker, E. M., & Gordon, E. W. (2025). Principles for assessment in the service of learning. 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.

Bennett, J. (2017, December 8). *Assessment FOR learning vs. assessment OF learning.* Pearson. https://www.pearsonassessments.com/professional-assessments/blog-webinars/blog/2017/12/assessment-for-learning-vs--assessment-of-learning.html

Chatterji, M. (2025). *User-centered assessment design: An integrated methodology for diverse populations.* The Guilford Press.

Darling-Hammond, L., & Adamson, F. (2014). *Beyond the bubble test: How performance assessments support 21st-century learning.* Jossey-Bass.

Darling-Hammond, L., Herman, J., Pellegrino, J. W., Abedi, J., Aber, J. L., Baker, E., Bennett, R., Gordon, E., Haertel, E., Hakuta, K., Ho, A., Linn, R. L., Pearson, P. D., Popham, W. J., Resnick, L., Schoenfeld, A. H., Shavelson, R., Shepard, L. A., Shulman, L., & Steele, C. M. (2013). *Criteria for high-quality assessment.* Stanford Center for Opportunity Policy in Education.

Foster, N., & Piacentini, M. (2025). Innovating assessment design to better measure and support learning. 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.

Gordon, E. W. (1995). Toward an equitable system of educational assessment. *The Journal of Negro Education, 64*(3), 360–372.

Gordon, E. W. (2020). Toward assessment in the service of learning. *Educational Measurement: Issues and Practice, 39*(3), 72–78.

Gordon, E. W. (2025). Series introduction: Toward assessment in the service of learning. 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.

Gordon, E. W., & Rajagopalan, K. (2016). *The Testing and Learning Revolution: The Future of Assessment in Education* (pp. 107–146). New York: Palgrave Macmillan US.

Huff, K. (2025). Designing and developing educational assessments for contemporary needs. 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.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15–21.

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.

Marion, S. F., Pellegrino, J. W., & Berman, A. I. (Eds.). (2024). *Reimagining balanced assessment systems.* National Academy of Education.

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

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment.* National Academies Press. https://doi.org/10.17226/10019

Penuel, W. R., & Watkins, D. A. (2019). Assessment to promote equity and epistemic justice: A use-case of a research-practice partnership in science education. *The Annals of the American Academy of Political and Social Science, 683*(1), 201–221. https://doi.org/10.1177/0002716219843249

Pellegrino, J. W. (2025). Arguments in support of innovating assessments. 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.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14.

Shepard, L. A. (2019). Classroom assessment to support teaching and learning. *The Annals of the American Academy of Political and Social Science, 683*(1), 183–200. https://doi.org/10.1177/0002716219843818

Shepard, L. A. (2021). Ambitious teaching and equitable assessment: A vision for prioritizing learning, not testing. *American Educator, 45*(3), 28–37, 48.

Tucker, E. M., & Armour-Thomas, E. (2025). Foundational Issues of Assessment in the Service of Learning. 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.