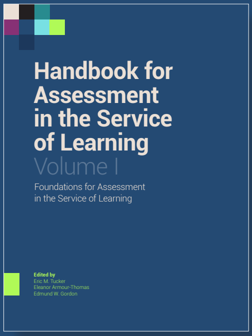
Principles for Assessment in the Service of Learning

Eva L. Baker, Howard T. Everson,
Eric M. Tucker, and Edmund W. Gordon

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:
Baker, E. L., Everson, H. T., 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.

Principles for Assessment in the Service of Learning

Eva L. Baker, Howard T. Everson, Eric M. Tucker, and Edmund W. Gordon

Abstract

This chapter offers a set of seven principles to guide the design and use of learning-focused assessments, that is, educational tests and assessments intended to support student learning. In the Handbook volumes, the principles were intended to assist chapter authors in considering these common elements in their contributions. In fact, some authors have chosen to include their own rendition of recommended principles. However, this brief chapter enumerates the principles developed by the participants of the Gordon Seminar. The principles are followed by a general rationale to support their inclusion as core components of an integrated design framework for assessments and concomitant learning and instruction. The seven principles are: (1) Assessment transparency; (2) Assessment focus and purpose; (3) Assessment support of learning processes (including attention, motivation, engagement, effort, and metacognition); (4) Assessment modeling of expectations and desired learning over time; (5) Assessment-linked instructional support including feedback; (6) Assessment equity with attention to learner variation; and (7) Assessment quality and validity involving the development of evidence to assure the value of assessment and learning to the learner and teacher.

Introduction

To guide the design, development, and use of assessments that are intended to support learning, we provide a set of principles to serve as the foundation for a conceptual framework for teachers, developers, and users of assessment results. Educational assessments with the deliberate intent of optimizing learning and teaching, as opposed to reporting student status, will have different characteristics from traditional large-scale assessments. Intended for use in any educational or training system, the proffered principles apply across academic disciplines and subject matter domains, settings, and ages and backgrounds of learners. The following principles form the crux of our recommendations.

- **Principle 1:** Assessment **transparency** provides clear information about assessment content and use to assist learners, teachers, administrators, and parents.
- **Principle 2:** Assessment **focus** is explicit and includes purposes, outcomes, progress indicators, and processes that can be transferred to other settings, situations, and conditions.
- **Principle 3:** Assessment design supports learners' **processes**, such as motivation, attention, engagement, effort, and metacognition.
- **Principle 4:** Assessments model the structure of **expectations** and **desired learning** over time.
- **Principle 5: Feedback**, adaptation, and other relevant instruction should be linked to assessment experiences.
- **Principle 6:** Assessment **equity** requires fairness in design of tasks and their adaptation to permit their use with respondents of different backgrounds, knowledge, and experiences.
- **Principle 7: Assessment quality and validity** should be available and reflect evidence related to assessment purpose to permit appropriate inferences and findings about quality, utility, and credibility.

Structure of Chapter

We provide the logic of the development of these principles in an overall rationale that includes the background of the Gordon Seminar for Assessment in the Service of Learning and the focus on assessment in support of learning. Next, we give an additional explanation relevant to the set of principles and briefly consider key topics in turn.

Finally, we offer a summary. Note that we have chosen not to include a significant number of footnotes or references within the text to allow the reader to make easy progress through this material, almost all of which will be expanded or illustrated in subsequent chapters. For those with interest in the scientific or experiential bases of the principles, we refer you to the selected bibliography at the conclusion of this narrative. For each principle, the selected bibliography provides a set of references that highlight its theoretical and empirical underpinnings.

A Rationale for Assessments Designed Explicitly to Inform Learning and Teaching

In the hands of skilled educators (e.g., teachers and developers), we believe that these principles will support a conceptual framework useful for designing educational assessments and tests to meet the needs of learners and of those engaged in helping people learn. The principles provide brief but concrete suggestions for repurposing and improving the design and use of educational assessments to enable students to improve their learning and performance in emerging educational contexts.

The background of these principles stemmed from the conversations of the Gordon Seminar: a weekly online meeting of scholars engaged in discussions for five years about how to follow up and operationalize the findings of *The Gordon Commission on the Future of Assessment in Education* (Gordon Commission on the Future of Assessment in Education, 2013) That publication was "created to consider the nature and content of American education during the 21st-century and how assessment can be used most effectively to advance that vision by serving the educational and informational needs of students, teachers, and society" (Pellegrino, 2014, p. 1).

The key message of the Gordon Commission underscored the importance of employing assessment to support the growth of learners in contrast to its dominant use to report student status in general, comparative formats. These status results almost always depend on annual or other periodic large-scale administrations of standardized tests. The test content of these examinations may be only loosely connected to the realities of classroom activities. For example, test developers with their administrative time constraints and psychometric requirements invariably choose to interpret curriculum standards or goals differently (and necessarily more generally) than individual classroom teachers.

In contrast to these uses of tests for status reporting practices, the Commission's Report emphasized that assessment and its findings should directly benefit student learning as well as mediate the efforts of teachers or other instructional designers to support their students. To be sure, scholars and practitioners have been advocating the use of assessment to support learning for decades, but the Commission pulled together then current views to explore the future. We believe that to be most useful, the learner should be given access to purposes, formats, content, contexts, and timing of assessments, as well as results and follow-up. Because the Seminar discussions fortuitously overlapped with the dramatic acceleration of advances in technology, such as artificial intelligence (AI) options, the promise of integration with assessment and learning became more feasible. The justification for our recommendations in the form of principles has many strands, but a common thread is clarifying what the assessments are about, attributes of their design, how they will be applied and scored, and how learning partners, that is, students and teachers, can benefit from their use far beyond the assessment itself. These principles overlap, with linked concepts to strengthen their effects. We now turn to the recommended principles.

Principle 1: Assessment transparency provides clear information about assessment content and use to assist learners, teachers, administrators, and parents.

Let's start with transparency. In practice, formal and classroom assessments are often presented either as surprises or with a level of secrecy prior to their use. As a result, students may be uncertain about why they are being assessed and on what topics and skills. Our notion is that the assessment content should be obvious to the students, not something they "psych out," or feel anxious or puzzled about. Transparent assessment practice should include sharing with the learners understandable models or guidelines for impending assessments. Transparent design means that the rules or parameters of the assessments (what's fair to be in them and what's fair to exclude) should be made clear for the boundaries of subject matter domains, as well as the types of thinking skills, assessment tasks, and criteria for evaluating the learners' performance. This clarity may suggest the enumeration of important objectives and content, or the use of graphics, such as maps or network models such as ontologies to show assessment components and their relationships. One may provide example questions and scoring guides in advance. Ideally, examples of content of interest, relevant cognitive demands, such as explanation or problem-solving, and task formats have occurred in relevant instruction. If not, efforts should be made to support learner preparation in advance of any assessment. Assessment transparency obviously aids the alignment of objectives, instruction, assessment, and effective teaching processes, a continuing requirement of coherent learning systems.

Principle 2: Assessment focus is explicit and includes purposes, outcomes, progress indicators, and processes that can be transferred to other settings, situations, and conditions.

The goal of understanding assessments is further supported by explication of the focus and purpose of the assessments in the second principle. Assessment focus refers to when in the learning process the assessment is positioned. Focus addresses whether the assessment attends to outcomes, progress in learning, or processes to be employed in the present program and then to be transferred to other situations or domains of knowledge. We identify both progress toward outcomes and processes to assist learning. Both domain-dependent and domain-independent processes are included. The purpose of assessment concerns its intended use. Is it for external reporting, grading, or consistent with our position, "assessment to support learning." The supportive use of assessment is fundamental to our principles; however, they may be applied to assessments for other purposes. In fact, it may be efficient for assessments to share more than one purpose, for instance, assessment to support ongoing learning as well as to be used in part for external reporting. Technology-supported environments can often provide procedures to aggregate responses obtained by individuals or groups during their learning activities to allow both uses.

Principle 3: Assessment design supports learners' processes, such as motivation, attention, engagement, effort, and metacognition.
Effective assessment design deliberately includes key student learning processes such as motivation, attention, engagement, effort, and metacognitive self-regulation. When assessments are structured to actively engage and enhance these processes (for example, by incorporating meaningful tasks, opportunities for reflection, and timely feedback), the assessment itself becomes a catalyst for learning and skill development. This principle also addresses processes included in Principle 2. Note that these processes may be domain-specific, such as close reading of literary texts or mathematics problem-solving approaches, or may be domain-independent, that is, applied to a variety of subject matters and tasks. However, domain-independent processes may have greater utility and range and apply across topics, content domains, situations, and time. For instance, processes involving metacognition (or self-analysis by learners of their own learning behaviors) can be used across the learners' lifespan of learning. Moreover, the process of learning itself can enable application of learning processes to other situations either closely or significantly varied from the conditions of learning. In psychology, this application is known as "transfer" and to be effective, the features or attributes to be applied to different situations or content need to be taught and adapted to these new tasks and settings. Types of these transferable skills are self-assessment, planning, managing attention, engagement, motivation, handling level of effort and avoiding distraction. Other transferable processes may involve interpersonal skills, for example, performing various roles in assessment that require peer collaborations.

Principle 4: Assessments model the structure of expectations and desired learning over time.
To consider further the learning and instructional interactions with assessment, Principle 4 highlights the importance of assessments to structure learners' understanding of important knowledge and skills. The format and expectations of questions and tasks on assessments can provide models for future learning. Some writers consider the practice effects of tests as a way to build student competencies. The form of questions and the requirements of answers are basic components here, although the overall content on an assessment can help students understand relationships within a topic or domain. Some assessments also give students the criteria in advance to guide evaluation. Although related to transparency, our emphasis here is about affecting learning. When the criteria are given to learners in advance of assessment, they can shape the nature and quality of their answers or more complex performances.

Principle 5: Feedback, adaptation, and other relevant instruction should be linked to assessment experiences.

Feedback, that is, direct follow-up on assessment performance, has long been an attribute of good instruction because it provides another set of strong cues about learning. Feedback varies in specificity or detail, which influences how difficult it is for students to infer the quality of their answer (and what to do about it). For instance, feedback can be given generally about the overall level of "goodness" of a response, such as an unelaborated grade on a paper. General feedback requires learners to infer (and perhaps err) about their own strengths and weaknesses. Offering a model paper or rubric for open-ended tasks adds opportunities for engagement as learners are expected to be able to identify similar and dissimilar features between their own constructions and the rubric or modeled responses. Providing elaborated explanations permits more guidance as feedback is crafted to distinguish among appropriate and somewhat lacking performance elements needed in the response. Typically, forms of feedback which give additional instruction add substantial time to the learning/assessment process. Peer feedback is another frequent way to engage all students in learning. When feedback is supplied by technology-rich systems rather than by the classroom teacher alone, it can not only support learners but also enable feedback to the instructors about their own success, allowing their review of summaries of student progress. This feedback to teachers affords them more guidance for subsequent instruction needed to support learners. Across the board, various forms of feedback have been found to be effective components of learning.

Principle 6: Assessment equity requires fairness in design of tasks and their adaptation to permit the use with respondents of different backgrounds, knowledge, and experiences.

Central to the concerns of the Gordon Seminar has been the principle of equity, a word which has many interpretations and which we hope to clarify. Equity considers attention to different experiences of learners, including their expectations for school, peer relationships, relevant knowledge, sense of belonging, and other differences in order to adjust assessment and learning to account for varied learner features. To some, equity has been understood to mean the erasure of such individual differences. That interpretation suggests that the outcomes would need to be the same for all groups of learners. In our vision of assessment and learning, we are impelled to give opportunities that connect to the diverse students we have. True equitable assessment may not result in comparable results, especially if learners have different levels of expertise in a domain or are best served by different task formats and support. Methods to consider in adapting assessments involve offering choices of topic or context, to allow collaborative efforts to support different strengths, or to give a range of resources in the assessment to shore up prior knowledge.

An additional requirement of learning-based assessment is the effort to adapt the assessment, instruction, and learning environments in order to adjust experiences to the students' particular strengths and growth areas. We say effort to adapt because plausible adaptations may not fit targeted students. Instructional adaptation is dependent on performance and often follows feedback. Generalizations about background and culture are subject to deliberate or implicit biases. Some attempts to adapt to diversity, such as homogeneous grouping, have drawn strong opposition by groups believing that adaptation is likely to keep students operating at lower levels of expertise. There is a range of evidence about the feasibility of adapting assessments and instruction to meet diverse learner needs. Nonetheless, feasibility alone does not guarantee enactment; time, training, and materials frequently constrain what teachers can do in real time. Assessment systems should therefore provide supports that anticipate common needs and help educators avoid systematic pitfalls for each student during learning. Adaptation is often made based on the answer to the most recent assessment task and usually only attends to content and skill levels instead of the far wider swath of individual differences. Again, giving students choices among topics or projects for learning would seem to offer one path, if we assume the choices appear to be real to students (piloting is an option here) and that students know how or have been taught to make good choices. We would expect in the future more precise adaptation will occur, adjusting to multiple variables including attention, motivation, and student preferences in addition to their prior performance. Learners should also be encouraged to approach individualized goals as well. At the heart of our conception of learning is the reality of human variance on a wide range of measured and unmeasured dimensions. Our focus on assessment and subsequent adaptation is an attempt to make the opportunities fair, and to advance and avoid systematic pitfalls for each student during learning.

Principle 7:Assessment quality and validity should be available and reflect evidence related to assessment purpose to permit appropriate inferences and findings about quality, utility, and credibility.

Although we rely on commercial purveyors of learning and assessment, it is obligatory to determine if there is evidence of the quality of the assessment and learning, before they are used widely with students. If assessments provide inaccurate or otherwise untrustworthy results, they undermine our ability to contribute to student learning and may give the students and teachers a mistaken view. Validity experts have created various models for use, some of which depend on many students making many responses, and analyses of complex statistical models. But validity does not need to depend on statistical complexities. The key point of validity is whether one can relate findings to the purpose and focus of the assessment; in particular, the evidence (e.g., student results) should closely relate to the desired inferences to be made about purposes and intended uses. Validity models may include components such as test content (sampling and distribution), response processes stimulated by the assessment and their agreement with the assessment's intent, and adequacy of scoring procedures to determine quality of performance. There are also equity elements that include the fairness of the process and content of the assessment. In assessments for learning, however, there is a tight interaction among the learning or instructional experiences, the assessment, and subsequent additional instruction, if needed. The quality of the assessment includes standard provisions, such as accuracy, representation, and framing so that learners can respond to appropriate queries. To investigate the empirical utility of learning-embedded assessments, one could imagine experimentally comparing learning sequences with and without the embedded assessments to determine the assessments' contribution to learning. Other elaborate approaches might compare all learners' response patterns with those of students deemed successful in the program or those of other experts. In all options, the evidence collected, for example, answers to questions or produced projects, closely relates to the desired inference about student learning. Validity also applies to categories into which students might fall. If the inferences from tests or assessments place students in competency groups (e.g., excellent, needs improvement, or more instruction), then the cut-scores or boundaries used to divide students into categories will also need attention. Note also that one cannot simply adapt validity models useful for one test purpose, such as accountability testing, to apply them to assessments for learning. Many efforts to date involve using both qualitative and quantitative data to draw inferences. Sadly, validity evidence is not often available for assessments embedded in learning systems.

Because the quality of the assessment may drive the entire process of learning and certification of students, the importance of validity or other data on assessment quality cannot be overstated. Since decisions may be made regarding external learning systems and assessments for use in schools and in homes, reports of validity and quality should be made available in common language. These nontechnical reports should detail the administration condtions, and the numbers and type of students from whom the data were collected to be sure there are essential matches with students for whom the systems are intended.

Summary

In this initial chapter we have deliberately taken a quick look at principles guiding our Gordon Seminar discussions that we have culled from a vast set of options that might have been written about assessment and learning. We noted that some of our contributors to the Handbook have created or referred to other sets of principles for assessment design and use. Of key interest to many scholars is the periodic publication of the *Standards for Educational and Psychological Testing* (AERA/APA/NCME, 2014), supported by three professional organizations, the American Educational Research Association, the American Psychological Association, and the National Council for Measurement in Education. These standards have tended to be conservative in that entries in them are made when there is consensus in the field. One can see the progress of concern for equity and fairness, first more extensively treated in 1999 and expanded in 2014. With the speed of technological innovations, such as the use of AI systems to clarify transparency, generate tasks measuring complex performance, evaluate student efforts, conduct validity studies, and prepare reports, we anticipate great change. Yet, we believe that the principles we have offered remain relevant to a wide range of contexts of learning-relevant assessments, such as individual assessments, assessments of groups or teams, and capturing students' development over time. Similarly, we believe the principles apply to assessments of content knowledge, skill development, affective and self-management processes, and zest for learning. In the subsequent chapters in these volumes, scholarly explanations and examples will be available to flesh out our initial descriptions of our principles.

References

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.

Gordon Commission on the Future of Assessment in Education. (2013). *To assess, to teach, to learn: A vision for the future of assessment* [Technical Report]. ETS. https://www.ets.org/Media/Research/pdf/gordon_commission_technical_report.pdf

Pellegrino, J. (2014). Assessment in the service of teaching and learning: Changes in practice enabled by recommended changes in policy. *Teachers College Record, 116* (110313). https://doi.org/10.1177/016146811411601102

Selected Bibliography

Assessment in the Service of Learning

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.

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

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

The Gordon Commission on the Future of Assessment in Education. (2013). *To assess, to teach, to learn: A vision for the future of assessment* [Technical Report]. https://www.ets.org/Media/Research/pdf/gordon_commission_technical_report.pdf

Pellegrino, J. (2014). Assessment in the service of teaching and learning: Changes in practice enabled by recommended changes in policy. *Teachers College Record, 116*(110313). https://doi.org/10.1177/016146811411601102

Ruiz-Primo, M. A., & Furtak, E. M. (2024). Classroom activity systems to supportambitious teaching and assessment. In S. F. Marion, J. W. Pellegrino, & A. I. Berman(Eds.), *Reimagining balanced assessment systems* (pp. 93–131). National Academy of Education.

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

Principle 1: Assessment transparency provides clear information about assessment content and use to assist learners, teachers, administrators, and parents.

Chung, G. K. W. K., Delacruz, G. C., Dionne, G. B., & Bewley, W. L. (2003). Linking assessment and instruction using ontologies. *Proceedings of the I/ITSEC, 25,* 1811–1822.

Clancey, W. J., & Shortliffe, E. H. (Eds.). (1984). *Readings in medical artificial intelligence: The first decade.* Addison Wesley. https://impact.dbmi.columbia.edu/~ehs7001/Clancey-Shortliffe-1984/Readings%20Book.htm

Gagné, R. M., & Briggs, L. J. (1974). *Principles of instructional design.* Holt, Rinehart & Winston.

Iseli, M. R., & Jha, R. (2016). Computational issues in modeling user behavior in serious games. In H. F. O'Neil, E. L. Baker, & R. S. Perez (Eds.), *Using games and simulations for teaching and assessment: Key issues* (pp. 21–40). Routledge.

Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment & Evaluation in Higher Education, 39*(7), 840–852. https://doi.org/10.1080/02602938.2013.875117

Moss, C. M., & Brookhart, S. M. (2012). *Learning targets: Helping students aim for understanding in today's lesson.* ASCD.

Principle 2: Assessment focus is explicit and includes purposes, outcomes, progress indicators, and processes that can be transferred to other settings, situations, and conditions.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives* (Complete edition). Longman.

Andrade, H. L., Bennett, R. E., & Cizek, G. J. (Eds.). (2019). *Handbook of formative assessment in the disciplines* (1st ed.). Routledge. https://doi.org/10.4324/9781315166933

Armour-Thomas, E., & Gordon, E. W. (2025). *Principles of dynamic pedagogy: An integrative model of curriculum instruction and assessment for prospective and in-service teachers.* Routledge.

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

Heritage, M. (2021). *Formative assessment: Making it happen in the classroom* (2nd ed.). Corwin.

Lee, C. D. (1998). Culturally responsive pedagogy and performance-based assessment. *The Journal of Negro Education, 67*(3), 268–279. https://www.jstor.org/stable/2668195?origin=crossref

van Merriënboer, J. J. G., & Kirschner, P. A. (2007). *Ten steps to complex learning: A systematic approach to four-component instructional design.* Routledge.

Principle 3: Assessment design supports learners' processes, such as motivation, attention,engagement, effort, and metacognition.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*(3), 261–271.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.). National Academy Press.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*(4), 438–481.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066x.34.10.906

Plass, J. L., & Kalyuga, S. (2019). Four ways of considering emotion in Cognitive Load Theory. *Educational Psychology Review, 31*(2), 339–359. https://doi.org/10.1007/s10648-019-09473-5

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Principle 4: Assessments model the structure of expectations and desired learning over time.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). Worth Publishers.

Black, P., Wilson, M., & Yao, S.-Y. (2011). Road maps for learning: A guide to the navigation of learning progressions. *Measurement: Interdisciplinary Research and Perspectives, 9*(2–3), 71–123.

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.

Gordon, E. G., & Bridgall B. L. (Eds.). (2006). *Affirmative development: Cultivating academic ability, critical issues in contemporary American education series.* Rowman & Littlefield Publishers, Inc.

Leonard, W. H., & Lowery, L. F. (1984). The effects of question types in textual reading upon retention of biology concepts. *Journal of Research in Science Teaching, 21*(4), 377–384. https://doi.org/10.1002/tea.3660210405

Phelps, R. P. (2012). The effects of testing on student achievement, 1910–2010. *International Journal of Testing, 12,* 21–43.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Principle 5: Feedback, adaptation, and other relevant instruction should be linked to assessment experiences.

Hattie, J. (2023). *Visible learning: The sequel: A synthesis of over 2,100 meta-analyses relating to achievement* (1st ed.). Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*(1), 79–97. https://doi.org/10.3102/00346543058001079

Linn, R. L. (1983). Testing and instruction: Links and distinctions. *Journal of Educational Measurement, 20*(2), 179–189.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology, 10,* Article 3087. https://doi.org/10.3389/fpsyg.2019.03087

Principle 6: Assessment equity requires fairness in design of tasks and their adaptation to permit the use with respondents of different backgrounds, knowledge, and experiences.

Armour-Thomas, E., McCallister, C., Boykin, A. W., & Gordon, E. W. (Eds.). (2019). *Human variance and assessment for learning.* Third World 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

Duran, R. P. (1989). Testing of linguistic minorities. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 573–587). Macmillan.

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

Herman, J. L., Bailey, A. L., & Martinez, J. F. (2023). Introduction to the special issue: Fairness in educational assessment and the next edition of the standards. *Educational Assessment, 28*(2), 65–67. https://doi.org/10.1080/10627197.2023.2215979

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

Oakes, J. (1986). Keeping track, part 1: The policy and practice of curriculum inequality. *Phi Delta Kappan, 68*(1), 12–17. https://www.jstor.org/stable/20403250

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

Sireci, S. G., & Randall, J. (2021). *Evolving notions of fairness in testing in the United States in the history of educational measurement.* Routledge.

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher, 32*(2), 3–13. https://doi.org/10.3102/0013189x032002003

Principle 7: Quality and validity should be available and reflect evidence related to assessment purpose to permit appropriate inferences and findings about quality, utility, and credibility.

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

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1–73.

Linn, R. L. (2010). Validity. In B. McGaw, P. L. Peterson, & E. L. Baker (Eds.), *International encyclopedia of education* (3rd ed., Vol. 4, pp. 181–185). Elsevier.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education/Macmillan.

Mislevy, R. J., Oliveri, M. E., Slomp, D., Crop Eared Wolf, A., & Elliot, N. (2025). An evidentiary-reasoning lens for socioculturally responsive assessment. In R. E. Bennett, L. Darling-Hammond, & A. Badrinarayan (Eds.), *Socioculturally responsive assessment: Implications for theory, measurement, and systems-level policy* (pp. 199–241). Routledge/Taylor & Francis.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*(1), 3–67.

Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychologist, 51*(1), 59–81.

Sireci, S. G. (2013). Agreeing on validity arguments. *Journal of Educational Measurement, 50*(1), 99–104.