Approaches to Assessment in Higher Education

Colin Madland

Last updated Sunday, Jul 17, 22:02 pm, PDT

Table of Contents

# A Review of the Literature on Assessment in Technology-mediated Higher Education

## Introduction

Assessing learning is a critical component of the work of higher education instructors, yet not enough is known about how instructors plan for and make decisions about how they will assess their students’ learning, particularly in the context of technology-mediated environments. In her important article, Shepard ([2000](#ref-shepardRoleAssessmentLearning2000)) argued at the turn of the century that contemporary (at the time) assessment structures originated in past models of curriculum and instruction which were popular in the early 1900s. These curricular models emphasized the work of psychologists like Thorndike ([Thorndike, 1905](#ref-thorndikeElementsPsychology1905)) and Skinner ([Skinner, 1938](#ref-skinnerBehaviourOrganisms1938)) who viewed the process of learning as being grounded in the mechanistic view of behaviourism where learning is the result of the precise and controlled input of ‘knowledge’ and reinforced with rewards for correct responses. As such, instructors (appropriately at the time) designed their assessments to align with the curricular goals of the time and assessed learning by determining whether or not a learner could provide the single correct response to a given question at a time removed from the instruction. However, in the latter half of the 20th century, when western psychologists discovered the ideas of Lev Vygotsky ([Vygotsky, 1978](#ref-vygotskyMindSociety1978)), a contemporary of the previously mentioned psychologists, curricula began to take a more social-constructivist approach that emphasized higher-order thinking, problem-solving in social contexts, and metacognitive skills over rote memorization. Unfortunately, it seemed that the efficiencies of testing memory, recognition, and recall through selected-response tests were too deeply embedded in the practices of HE instructors who resisted changing their assessments to match the new curricular goals. Barnett ([2007](#ref-barnettWillLearnBeing2007)) contends that these new goals lead to a state of *supercomplexity* with which both learners and instructors must grapple, and which Garrison et al. ([2015](#Xd550762775e92860104485be5231f035aab8d86)) argue requires deep changes to how instructors approach teaching and how learners approach learning.

Shepard argues for the need to integrate assessment and instruction in such a way as to engage learners in authentic performance tasks more suited to modern understandings of cognition. It appears now that, in the twenty-two years since Shepard wrote her paper, the goals of early 21st century curricula have continued to diverge from those of the 20th century with the World Economic Forum identifying competencies in collaboration, analytical thinking, creative problem-solving, and continual learning as being priorities for 21st century employers ([Forum, 2020](#Xaf0f34faf6db1662e8da0b9850d3b9cfe5dbeed)). Additionally, digital and networked technologies have become ubiquitous in society and higher education ([Broadfoot, 2016](#X51a879b1b801bbc5ad6e7fc588f4456f2973b2d); [Pellegrino & Quellmalz, 2010](#Xd3220b3c4c74762840488f1ab7714e7e895bc2e); [Webb & Ifenthaler, 2018](#ref-webbAssessmentTwentyFirstCentury2018)) and the influence of digital technologies on instructors approaches to assessment remains under researched. Models of assessment which prioritize testing skills in a manner aligned with 20th century curricular models are no longer adequate because they no longer align with the priorities of modern HE ([Broadfoot, 2016](#X51a879b1b801bbc5ad6e7fc588f4456f2973b2d); [Crooks, 1988](#ref-crooksImpactClassroomEvaluation1988); [Pellegrino & Quellmalz, 2010](#Xd3220b3c4c74762840488f1ab7714e7e895bc2e); [Timmis et al., 2016](#X9783eef1c37e492f1192ce9092108eb526b7f44)).

Despite Shepard’s exhortation to integrate modern curricular goals with aligned assessment practices, both local and systemic change in this area has been elusive ([Broadfoot, 2016](#X51a879b1b801bbc5ad6e7fc588f4456f2973b2d); [Earl, 2013](#ref-earlAssessmentLearningUsing2013)). This is in part because the approaches that individual instructors take to assessment are driven by complex factors at the individual level as well as local and systemic levels ([Black & Wiliam, 1998](#ref-blackAssessmentClassroomLearning1998); [DeLuca, Coombs, et al., 2019](#ref-delucaDifferentialSituatedView2019); [Willis et al., 2013](#X731be6e2e0895c79e0f034fc829fd623ca4b0d3)). The variety of factors, combined with the relative freedom instructors have to design their courses and assessments leads to approaches to assessment that are variable, idiosyncratic, and often influenced by individual instructors’ past experiences rather than a deep understanding of assessment theory ([Lipnevich et al., 2020](#ref-lipnevichWhatGradesMean2020); [Massey et al., 2020](#ref-masseyAssessmentLiteracyCollege2020)).

{–Researchers are engaged in important work exploring how K12 teachers approach assessment in their classrooms, including DeLuca et al.’s ([2016](#Xedee9bc10fe7ddda531b14b05542852ae9f5f1e)) *Approaches to Classroom Assessment Inventory*, but there has been less work in relation to technology-mediated higher education. –}

Given the confluence of the preceding factors, the purpose of this literature review will be to synthesize and analyze the literature related to approaches to assessment among higher education instructors. The review will begin with a survey of historical conceptions of assessment, followed by a deeper analysis of the literature since 2010 with a focus on assessment in technologically-mediated higher education.

This review is guided by four research questions:

1. What factors shape higher education instructors’ approaches to assessment?
2. What are the major themes or patterns in the literature related to approaches to assessment in higher education?
3. What are the major themes or patterns in the literature related to the impact of technology on assessment in higher education?
4. What gaps exist in the literature related to approaches to assessment in technology-mediated higher education?

## A Brief History of Classroom Assessment Research

There are deep and rich bodies of literature addressing educational assessment *writ large*, and it would not be possible in the scope of this article to adequately cover the full breadth and depth of the published research ([Joughin, 2009](#X2997c928acaf5e247ad3a813d62dfa653f9bffa)). Consequently, while any starting point may seem somewhat arbitrary, the author chose to begin with the genesis of modern thinking of assessment being divided into “summative assessment” (for the purposes of providing a final judgement of learning) and “formative assessment” (for the purposes of providing feedback to learners and instructors to improve learning). This framing of assessment was popularized by Benjamin Bloom and remains prominent in the literature.

Among the more influential publications related to modern views of assessment (then usually called “evaluation”) was Scriven’s ([1967](#ref-scrivenMethodologyEvaluation1967)) article in which he drew distinctions between “formative” and “summative” evaluation. Formative evaluation was described as evaluation for the purposes of improvement, and summative evaluation was seen as a validation of the quality of work at the end of a process. This distinction was quickly incorporated into Bloom’s ([1968](#ref-bloomLearningMasteryInstruction1968)) ideas related to mastery learning and began to be promoted as a model for educational reform. However, by the late 1990s, when Black and Wiliam ([1998](#ref-blackAssessmentClassroomLearning1998)) published their landmark review of the literature on assessment, the idea of formative assessment was still not well-defined or implemented. Black and Wiliam framed formative assessment as “encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged” ([1998, pp. 7–8](#ref-blackAssessmentClassroomLearning1998)). Although Black and Wiliam came to very strongly-stated conclusions about the value of formative assessments (e.g. “The research reported here shows conclusively that formative assessment does improve learning. The gains in achievement appear to be quite considerable, and as noted earlier, amongst the largest ever reported for educational interventions.” ([1998, p. 61](#ref-blackAssessmentClassroomLearning1998))), reliance on summative assessments in HE remained high ([Harlen & Deakin Crick, 2002](#ref-harlenSystematicReviewImpact2002); [Hébert, 2021](#ref-hebertOnlineRemoteProctoring2021); [Lipnevich et al., 2020](#ref-lipnevichWhatGradesMean2020)).

Also towards the end of the 20th century, the idea of *assessment literacy* was introduced by Richard Stiggins([1991](#ref-stigginsAssessmentLiteracy1991)). Stiggins argued that, in K12 educational contexts, society had come to expect high educational attainment, but teacher education programs did little or nothing to prepare graduates to be able to certify that their learners were actually meeting targets. Stiggins blamed this in part on the fact that in the United States, there was no specific requirement for teachers to be trained in assessment, and consequently, teachers, administrators, and those who train them were essentially assessment illiterate. Stiggins described assessment literate educators as those who “have a basic understanding of the meaning of high- and low quality assessment and are able to apply that knowledge to various measures of student achievement” (p. 535).

An intuitive reading of this history suggests that the decade of the 1990s and into the early 2000s may have been pivotal in determining the direction of the discourse in assessment for the following 10-15 years. The 1990s were bookended by Stiggins’ article in 1991 and Black and Wiliam’s review in 1998. The emphasis on professional standards and skill-based assessment literacy, in alignment with Stiggins’ article, dominated teacher education programs in the 1990s and for the 20 years following. However, Black and Wiliams’ 1998 review of the literature and recommendations promoted a significantly greater emphasis on formative assessment ([2011](#X36e21314cdda5af289172df8b2fdcbf83ba9ea4)). The following sections trace the development of assessment literacy and then the published literature on formative assessment. ### Professional Standards / Assessment Literacy

The recognition of assessment literacy as a critical competency for educators was influenced by the growing demands for teacher and school accountability in the post-WWII era in the USA and Canada, particularly the *Elementary and Secondary Education Act* (ESEA), passed in 1965, and the *No Child Left Behind* (NCLB) act, passed in 2002 (see [DeLuca, 2012](#ref-delucaPreparingTeachersAge2012) for a detailed discussion). As such, conceptualizations of assessment literacy tended to be based on sets of standards to which K-12 teachers were obligated. Table 1 provides a brief overview of each of the sets of standards published between 1990 and 2015, including Brookhart’s ([2011](#X36e21314cdda5af289172df8b2fdcbf83ba9ea4)) critique of the 1990 *Standards* as being too focused on 20th century skills, rather than formative assessment practices and 21st century pedagogies. See Appendix A for a full listing of each of the standards and their constituent skills or characteristics.

**Table 1**

*Assessment Standards Compared*

| Title/Article Citation | Description |
| --- | --- |
| Standards for Teacher Competence in Educational Assessment of Students ([AFT et al., 1990](#ref-aftStandardsTeacherCompetence1990)) | The initial set of standards, published by a committee of representatives from the American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association. The *Standards* are a list of seven skills expected of teachers. |
| Principles for Fair Student Assessment Practices for Education in Canada ([Committee, 1993](#Xfa1f130840b43bd3aaba7652d3b73b050d23ae4)) | A group of Canadian educators published the *Principles for Fair Student Assessment Practices for Education in Canada*, Part A of which was a list of 37 guidelines related to five principles of fair student classroom assessment and was based on the 1990 *Standards* (see Appendix A). Part B was focused on externally-developed standardized tests. |
| Assessment Literacy For the 21st Century ([Stiggins, 1995](#ref-stigginsAssessmentLiteracy21st1995)) | Stiggins outlined five characteristics of sound assessments. |
| Educational Assessment Knowledge and Skills for Teachers ([Brookhart, 2011](#X36e21314cdda5af289172df8b2fdcbf83ba9ea4)) | Twenty years following the publication of the *Standards*, Brookhart ([2011](#X36e21314cdda5af289172df8b2fdcbf83ba9ea4)) argued that the 1990 *Standards* had become outdated because they did not address either the growing practices and ideas of formative assessment (assessment *for* and *as* learning) or standards-based assessment and that they needed to be revised. Brookhart suggested a list of 11 skills (see Appendix A) to adjust the focus of the 1990 *Standards* to be in greater alignment with more modern conceptions of assessment. |
| Classroom Assessment Standards for PreK-12 Teachers ([Klinger et al., 2015](#X29739ab7fce6a0c61ad93e769577d0974c2bf34)) | Finally, in 2015, the Joint Committee on Standards for Educational Evaluation (JCSEE), with key representatives from both Canada and the USA, published the most recent set of standards, called the *Classroom Assessment Standards for PreK-12 Teachers* (see Appendix A). The JCSEE standards are grouped into three broad domains (foundations, use, and quality), each with five or six related standards. |

The policy environment in the United States combined with assessment research and the publication of various sets of standards as highlighted above served to promote an emphasis on measuring outcomes using standardized tests and using the results from those tests as data points for evaluating teacher and administrator competence ([Brookhart, 2011](#X36e21314cdda5af289172df8b2fdcbf83ba9ea4)).  
### Formative Assessment {-}

Following Black and Wiliam’s ([1998](#ref-blackAssessmentClassroomLearning1998)) review, the National Research Council’s (NRC) 2001 foundational report *Knowing What Students Know: The Science and Design of Educational Assessment*, (hereafter, the *NRC report*) advanced understanding of assessment with their definition of assessment as “reasoning from evidence” ([2001, p. 43](#ref-pellegrinoKnowingWhatStudents2001)), or, more specifically “a process of drawing reasonable inferences about what students know on the basis of evidence derived from observations of what they say, do, or make in selected situations” ([Pellegrino et al., 2001, p. 112](#ref-pellegrinoKnowingWhatStudents2001)), based on Mislevy’s ([1994, p. 4](#ref-mislevyTestTheoryReconcieved1994)) assertion that “test theory is machinery for reasoning from students’ behavior to conjectures about their competence, as framed in a particular conception of competence.” Such a parsimonious description, however, may hide some of the complexities of fairly and equitably coming to know what learners know and can do in relation to particular outcomes. Since knowledge of a particular domain cannot be directly observed in a learner, and therefore cannot be quantified, instructors must rely on data gathered during the teaching process to support a particular inference about what a learner *probably* knows. The data gathered from performance tasks such as exams, essays, portfolios, labs, etc, become *evidence* when they support an inference about what a learner knows and can do ([Black & Wiliam, 2018](#ref-blackClassroomAssessmentPedagogy2018)).

The *NRC* report, which Black played a role in creating, used Black and Wiliam’s review quite prominently throughout in building the argument for formative assessment as opposed to a primary reliance on summative testing (see pp. 31, 38, 41, 226, 235, and 236). Pellegrino et al. ([2001](#ref-pellegrinoKnowingWhatStudents2001)) highlight that there are multiple purposes for assessment, assessment to assist learning, assessment of individual acheivement, and assessment to evaluate programs. They argue that it is important to recognize that not all assessments fit with all purposes of assessment, and that the more purposes an assessment is intended to serve, the less likely it is to serve all purposes well. Black and Wiliam note that these multiple purposes create tension for teachers who must navigate and align diverse purposes. However, as some jurisdictions have prioritized summative assessment structures to fulfil the purposes of assessing individual achievement and evaluating programs, it is often the case that the purpose of serving the learning needs of learners is sacrificed.

In the first decade of the 21st, researchers increasingly recognized the complexity of assessment practices and the factors which drive them. Brown ([2004](#Xbf36e976fd90f94421b02cbc8fcdfde6dafc0f9)) showed in a study of K12 teachers in New Zealand that teachers’ “conceptions of assessment” (p. 302), or the organizing frameworks teachers use to understand the phenomenon of assessment, are not simple constructs, but are complex and interconnected. Brown’s model showed four common conceptions of assessment among K12 teachers:

* improvement of teaching and learning,
* school accountability,
* student accountability, or
* treating assessment as irrelevant.

While Brown’s model nods at formative assessment in mentioning improvement in teaching and learning, there remained a strong emphasis on accountability for learners, teachers, and schools. However, as thinking progressed, James ([2008](#ref-jamesAssessmentLearning2008)) highlighted important connections between assessment practices and learning theory, noting that both behaviourism and cognitivism emphasize the acquisition of knowledge as a sort of commodity and thus both lead to assessment strategies that emphasize the individual reproduction of knowlege that has either been taught, or has been constructed in meaning-making exercises. Contrary to these approaches, socio-cultural learning theory emphasizes the communal production of meaning through social tools and social interactions between peers and more capable ‘others’ ([Vygotsky, 1978](#ref-vygotskyMindSociety1978)). Thus, assessment in socio-cultural contexts, which James argues is under-theorized, ought to be holistic and situated within the communal learning process, rather than an event that happens sometime after the learning has happened. Similarly, Fletcher et al. ([2012](#Xa55a5d631b402fa2c3fda6e3997bc5d48ccd027)) note that teachers who view teaching as the transmission of knowledge tend to rely more heavily on assessment tasks which require learners to reproduce what they have been taught. On the other hand, teachers who take a more socio-constructivist view that knowledge is built over time tend to conceive of assessment as being an integral part of the teaching and learning process as opposed to a culminating activity.

The second decade of the 21st century marked a further turn towards more complex theories of assessment and an emphasis on formative assessment. Brookhart ([2011](#X36e21314cdda5af289172df8b2fdcbf83ba9ea4)) published her influential article criticizing the overemphasis on dated forms of assessment and proposed an updated set of eleven standards which included much greater emphasis on formative assessment skills. Earl ([2013](#ref-earlAssessmentLearningUsing2013)) commented that the purposes of assessment might sometimes conflict with each other, such as when assessment might be used for both formative feedback and also as evidence supporting a summative grade. Earl contends that the purposes of assessment can be categorized as assessment *of* learning (summative judgements of learner knowledge), assessment *for* learning (formative feedback loops guiding and directing future learning), and assessment *as* learning (learners using metacognitive skills to self-regulate in learning contexts). While Earl’s taxonomy is primarily focused on how instructors and learners use assessment data, others note that conceptions of assessment extend well beyond this. DeLuca et al.([2013](#X32b95a94fb3076f8d0c577bd9276160db2e95e5)), for example, proposed that there are four predominant conceptions of assessment among teachers, namely, “(a) assessment as testing, (b) assessment as format, (c) assessment as purpose and (d) assessment as process” (p. 110). They argue that their taxonomy represents an ordered progression of complexity beginning with a primary focus on summative testing and progressing through to a conception of assessment that recognizes the complexity of integrated teaching and learning processes.

The most recent set of standards, published by the Joint Committee on Standards for Educational Evaluation ([Klinger et al., 2015](#X29739ab7fce6a0c61ad93e769577d0974c2bf34)) represents a further break from the former mindset of assessment as measurement and focused on three areas. First is the foundational knowledge needed to understand the purposes of assessment and how to design assessments that allow learners to demonstrate their ability and to communicate results with stakeholders. The second area of required competence is in the use of assessment data to analyze learner performance and provide formative feedback to improve future performance. The third area has to to with the quality of assessments such that they are culturally and linguistically appropriate, account for the differential needs of all learners, that they are unbiased, reliable, and support valid interpretations of assessment data and finally that they are continually revised and updated to improve overall quality.

This brief overview of various sets of standards, while incomplete, provides a picture of how researchers studying assessment have come to view assessment as a much more holistic and whole-person endeavour for both teachers and learners compared to early views emphasizing measurement grounded in behaviourist pedagogies.

### 0.0.1 Approaches to Assessment

In the context of this progression towards more formative assessment practices, other researchers began exploring the complexities of assessment by considering the many factors that comprise individual teachers’ approaches to assessment. DeLuca et al. ([2016](#X773f46dba91e2266ea33fc285a547bc86776461)) used their earlier framework of conceptions of assessment ([2013](#X32b95a94fb3076f8d0c577bd9276160db2e95e5)) and proposed the idea of *approaches to assessment*, representing an intentional move away from basing assessment practice on skills and competencies reified in sets of standards ([DeLuca et al., 2016](#Xedee9bc10fe7ddda531b14b05542852ae9f5f1e)). DeLuca et al. ([2016](#X773f46dba91e2266ea33fc285a547bc86776461)), citing ([Newton, 2007](#Xf29b30932ade5cdbd28a39c6700b1e19e97b792)) argue that the varying conceptions of assessment as well as the large number of factors that shape teachers’ approaches to assessment, all of which can be contradictory, created confusion and possibly hindered the progressive development of teachers’ competency in assessment. The idea of *approaches to assessment* is grounded in the complexity of assessment as well as researchers’ recognition that skill-based sets of standards fail to account for all of the factors shaping assessment practice in modern schools.

DeLuca et al.’s model ([2016](#Xedee9bc10fe7ddda531b14b05542852ae9f5f1e)), *approaches to classroom assessment* could be seen to align with the *assessment* component of Biggs’ 3P model of teaching and learning ([1999](#ref-biggsWhatStudentDoes1999); [1993](#ref-biggsTheoryPracticeCognitive1993)) and so aligns well with a concise model of teaching and learning in higher education. The *Approaches to Classroom Assessment* model is based on the JCSEE standards ([Klinger et al., 2015](#X29739ab7fce6a0c61ad93e769577d0974c2bf34)) and describes four themes of AL, each with three dimensions. The model represents somewhat of a break from previous models in that it references *approaches to assessment* rather than *assessment literacies*. This is a reflection of the authors’ view that language around *literacies* and *competencies* may indicate a reliance on “correct” views or methods rather than the complex array of influences that lead to multiple legitimate approaches as identified in the literature ([DeLuca, Coombs, et al., 2019](#ref-delucaDifferentialSituatedView2019); [Willis et al., 2013](#X731be6e2e0895c79e0f034fc829fd623ca4b0d3)).

The themes DeLuca et al ([2021, p. 10](#Xcffd07f338ae9c1e0f3a4638f84c8ddd08fd0be)) describe along with their associated dimensions are listed below and illustrated in figure 2:

1. Assessment purposes.
   * Assessment *of* learning
     + Teachers’ use of evidence to summate student learning and assign a grade in relation to students’ achievement of learning objectives
   * Assessment *for* learning
     + Teachers’ and students’ use of evidence to provide feedback on progress towards learning objectives (i.e., inform next steps for learning and instructions).
     + Involves both teacher-directed and student-centred approaches to formative assessment.
   * Assessment *as* learning
     + Focuses on how the student is learning by providing feedback or experiences that foster students’ metacognitive abilities and learning skills (e.g., self-assessment, goal-setting, learning plans).
     + Involves teachers but is primarily student-centred.
2. Assessment process
   * Design
     + Focuses on the development of reliable assessments and items that measure student learning in relation to learning objectives.
   * Use/scoring
     + Focuses of the adjustment and use of scoring protocols and grading schemes to respond to assessment scenarios.
   * Communication
     + focuses on the interpretation of assessment results and feedback through communication to students and parents.
3. Assessment fairness
   * Standard
     + Maintains the equal assessment protocols for all students.
   * Equitable
     + Differentiates assessment protocols for formally identified students (i.e., special education or English language learners)
   * Differentiated
     + Individualizes learning opportunities and assessments that address each student’s unique learning needs and goals.
4. Assessment theory
   * Consistent
     + Works to ensure consistency in results within assessments, across time periods, and between teachers.
   * Contextual
     + Works to ensure assessment or evaluation measures what it claims to measure (i.e., learning objectives) and promote valid interpretations of results.
   * Balanced
     + Works to ensure consistency in measuring what an assessment or evaluation intends to measure, and degree to which an assessment or evaluation measures what it claims to measure.

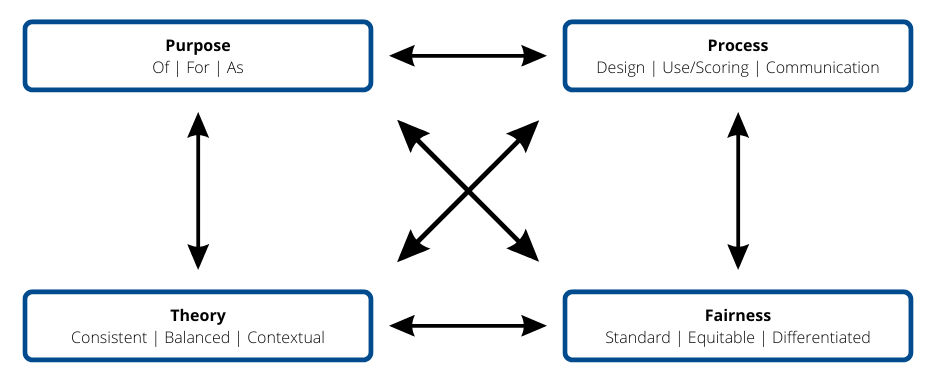


Figure 1. Approaches to Assessment ([DeLuca et al., 2021](#Xcffd07f338ae9c1e0f3a4638f84c8ddd08fd0be))

DeLuca et al. ([2016](#Xedee9bc10fe7ddda531b14b05542852ae9f5f1e), p 252) initially proposed three general approaches to classroom assessment aligning with the four themes listed above.

Table 2. Assessment Literacy Themes with Associated Approaches to Assessment

|  | Approaches to Assessment |  |  |
| --- | --- | --- | --- |
| Assessement Literacy Theme | Approach A | Approach B | Approach C |
| Assessment Purposes | Assessment of Learning | Assessment for Learning | Assessment as Learning |
| Assessment Processes | Design | Use/Scoring | Communication |
| Assessment Fairness | Standard treatment | Equitable treatment | Differentiated approach |
| Measurement Theory | Reliability | Validity | Reliability/Validity |

## Factors

In exploring higher education instructors’ approaches to assessment, it is helpful to look to research in the K12 sector as the idea of approaches to assessment originated in the context of K12 teaching. There have long been concerns that K12 teachers do not graduate with sufficient knowledge of assessment, even though they complete undergraduate degrees which have come to include specific courses in assessment ([Stiggins, 1991](#ref-stigginsAssessmentLiteracy1991)). Given this, in addition to the fact that higher education instructors generally do not encounter any specific pedagogical or assessment preparation in their graduate degrees, we can infer that the assessment capabilities of higher education instructors are much lower than those of K12 teachers. This is supported in research conducted by Massey et al. ([2020](#ref-masseyAssessmentLiteracyCollege2020)) who investigated the impact of a semester-long instructional development course on the conceptions of assessment and assessment confidence of new faculty. Pre-course responses indicated a predominant view that assessment is primarily about summative testing to confirm that learners have retained information from didactic learning experiences. They found that the instructional development course encouraged faculty to deepen their understandings of assessment, and as they did so, their conceptions of assessment became more nuanced and complex. At the same time, following the increased depth in conceptions, researchers noted that participants’ confidence in assessment also increased.

Popham ([2009](#ref-pophamAssessmentLiteracyTeachers2009)) describes two often competing types of assessments, *classroom* assessments and *accountability* assessments. Classroom assessments are typically teacher-created, although some teachers use publisher-created materials such as test banks, and are used to assign grades to learners or to inform future instruction. Accountability assessments are those which are usually government-imposed, standardized tests, most often deployed in K12 settings, although there are notable higher education contexts where such accountability tests are prominent, namely programs which prepare learners for external regulatory exams (e.g. nursing, accounting). Popham argues that the relative emphasis on either of these types of assessment will impact the decisions teachers make in their instructional and assessment approaches.

Offerdahl and Tomanek ([2011](#X568de4367a2adb3cc7734bdfe3ece898df67951)) investigated instructors’ assessment thinking in relation to being presented with new assessment strategies. They found that changes in assessment *practice* didn’t necessarily follow from changes in assessment *thinking* and that there are many factors which influence both assessment thinking and practice.

Willis et al. ([2013](#X731be6e2e0895c79e0f034fc829fd623ca4b0d3)) highlight the importance of understanding assessment literacies (plural) as being understood in relation to learning theory and that assessment is an “ethical practice that is social, dynamic, and layered” (p. 242). They situate their discussion in the idea of “horizontal” and “vertical” discourses as articulated by Bernstein ([1999](#X4d78eac5dad262ef2a233c8ad9b66b0f5b3cd1c)). A horizontal discourse is characterized as being common sense knowledge that is presumed to be known by local communities. It is context dependent and can be contradictory across communities while remaining internally coherent within communities. Bernstein contrasts this with the idea of vertical discourses, which are characteristically organized in a hierarchy with explicit principles which are systematically enacted (such as in the sciences) or rely on “a series of specialised languages with specialised modes of interrogation and specialised criteria for the production and circulation of texts” ([Bernstein, 1999, p. 159](#X4d78eac5dad262ef2a233c8ad9b66b0f5b3cd1c)), as in the social sciences. As vertical discourses rely on specialized language and expertise, they must be “decoded or translated (pedagogised)” ([Singh, 2002, p. 4](#Xed165d3d5256ba50c130dd813ca4571cd043257)) in order for those outside the domain to access that particular knowledge. Willis et al. ([2013](#X731be6e2e0895c79e0f034fc829fd623ca4b0d3)) argue that, since teachers are influenced by both horizontal and vertical discourses, they must constantly work to balance vertical and horizontal discourses. They use the example of the vertical discourse reflected in the “measurement conception of assessment” ([2013, p. 244](#X731be6e2e0895c79e0f034fc829fd623ca4b0d3)) in which assessment is viewed in light of positivist scientific paradigms, behaviourist learning theory, and an emphasis on testing learning at a time removed from instruction. This discourse is seen prominently in Confucian traditions ([Carless, 2011](#ref-carlessTestingProductiveStudent2011)), as well as cultures that use high-stakes testing for learner advancement and teacher accountability, such as the USA ([Popham, 2009](#ref-pophamAssessmentLiteracyTeachers2009)). Contrary to this prominent vertical discourse is the horizontal discourse grounded in socio-cultural perspectives exemplified in local cultures which downplay the role of high-stakes testing and instead focus on assessment as an integral component of learning ([DeLuca & Klinger, 2010](#Xba6256bc1df8223d2235ef500b76f76afb0417a)).

Looney et al. ([2017](#X2f43bdc2c6d6b8b53de1e6df50a2037ee9c597a)) argue that teacher assessment literacy is too instrumental a focus and that *conceptions of assessment* tend to neglect affective feelings related to teacher self-efficacy. They argue that teacher assessment identity encompasses these factors and that there are significant ontological dimensions to assessment. Teachers assess the way they do because of *who* they are in addition to what they know and are able to do.

([DeLuca, Willis, et al., 2019](#ref-delucaPoliciesProgramsPractices2019))

### Assessment Cultures

([Massey et al., 2020](#ref-masseyAssessmentLiteracyCollege2020))

([Birenbaum, 2014](#Xd1039a152831aefb97ded2531aec686e4bcf52b)) - assessment for learning cycle corresponds to inquiry sycle

([DeLuca et al., 2021](#Xcffd07f338ae9c1e0f3a4638f84c8ddd08fd0be)) ([Birenbaum, 2016](#Xcfce0211190b51c59aed4ad2e41b47cb53c6a86))

[delucaCrossculturalComparisonGerman2020] [willisConceptualisingTeachersAssessment2013] ([Hébert, 2021](#ref-hebertOnlineRemoteProctoring2021))

### 0.0.2 Impact Of Technology On Approaches To Assessment In Higher Education

While digital technologies have become ubiquitous in higher education ([Broadfoot, 2016](#X51a879b1b801bbc5ad6e7fc588f4456f2973b2d)), research suggests that the impact of technology on pedagogy, including assessment practices, in higher education has not kept up with the potential. Garrison et al. ([2015](#Xd550762775e92860104485be5231f035aab8d86)) introduce the idea of *shared metacognition* as being a necessary competency for learners in 21st century higher education. They build a case that modern learners must contend with *supercomplexity* ([Barnett, 2007](#ref-barnettWillLearnBeing2007)), which refers to the open-endedness and ambiguity characteristic of modern society and culture. Supercomplexity is both compounded and mitigated through the integration of digital technologies in higher education. While there is additional cognitive load associated with learning to learn through technology, it is also true that technology can help learners organize information through *cognitive offloading* ([Risko & Gilbert, 2016](#ref-riskoCognitiveOffloading2016)). Supercomplexity requires learners to engage in both cognitive offloading and shared metacognition in order to be successful, especially in light of the goals of modern universities.

Garrison et al. employ the same argument as Shepard ([2000](#ref-shepardRoleAssessmentLearning2000)) when they assert that modern technology-mediated learning environments tend to be grounded in out-dated approaches.

Some researchers use terms, such as “technology-*enhanced*”, “technology-*enabled*, or”technology-*rich*” ([Eyal, 2012, p. 37](#ref-eyalDigitalAssessmentLiteracy2012)) assessment, which show a positivity bias towards the use of technoloogy in higher education. Although this terminology will be a component of the search process, this review will use the more neutral term “technology-mediated assessment” whenever possible in light of the fact that assessment is not always “enhanced”, “enabled”, or “enriched” with the use of technology.

Modern universities are under pressure from public and private funding agencies ([Hébert, 2021](#ref-hebertOnlineRemoteProctoring2021)) and employers to demonstrate that graduates are equipped for the demands of citizenship in the 21st century ([Pellegrino & Quellmalz, 2010](#Xd3220b3c4c74762840488f1ab7714e7e895bc2e)). Technological changes in society have impacted how people live, work, play, and learn in many ways leading to additional pressures on universities to respond to new realities ([Forum, 2020](#Xaf0f34faf6db1662e8da0b9850d3b9cfe5dbeed)). As a result, universities have incorporated many technologies into how they operate, including student information systems, faculty career tracking systems, and learning management systems, to name a few. Technologies have also impacted how instructors teach, with many instructors incorporating digital tools such as the aforementioned learning management systems, but also in-class slide-decks to accompany lectures (often replacing older technologies, such as overhead projectors and chalk boards), digital response systems, digital distribution and gathering of documents, digital feedback, networked learning environments (i.e., blogs, git-based repositories, wikis, and other collaborative digital learning environments), and, more recently, artificially intelligent agents and algorithms used to interact with learners and even evaluate learner artifacts. Many of these technologies have allowed both universities and instructors to automate, and therefore scale up, processes and procedures that formerly consumed significant time and labour, however, in most cases, they have not fundamentally changed the kind of work that is being done ([Broadfoot, 2016](#X51a879b1b801bbc5ad6e7fc588f4456f2973b2d)). For example, automated grading of selected-response tests using a learning management system or a bubble sheet has greatly reduced the amount of time it takes to score selected-response tests, saving instructors significant time. However, this technology has not fundamentally changed the selected-response test itself. Similarly, collecting digital artifacts, like essays, has improved tracking and likely reduced the number of lost essays, but it has not fundamentally changed the nature of the assessment task. Despite the widespread adoption of technologies for many tasks in higher education, it would seem that technology has not yet significantly transformed how instructors encourage or assess learning in their classes ([Garrison & Akyol, 2015](#Xd550762775e92860104485be5231f035aab8d86)).

# Appendix A

Select Standards for Educational Assessment

## 0.1 Standards for Teacher Competence in Educational Assessment of Students

1. Teachers should be skilled in *choosing* assessment methods appropriate for instructional decisions.
2. Teachers should be skilled in *developing* assessment methods appropriate for instructional decisions.
3. The teacher should be skilled in administering, scoring, and interpreting the results of both externally-produced and teacher-produced assessment methods.
4. Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.
5. Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments.
6. Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.
7. Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

National Council on Measurement in Education, American Federation of Teachers, & National Education Association. (1990). Standards for Teacher Competence in Educational Assessment of Students. <https://eric.ed.gov/?id=ED323186>

## 0.2 Principles for Fair Student Assessment Practices for Education in Canada

### 0.2.1 Developing and Choosing Methods for Assessment

1. Assessment methods should be developed or chosen so that inferences drawn about the knowledge, skills, attitudes, and behaviors possessed by each student are valid and not open to misinterpretation.
2. Assessment methods should be clearly related to the goals and objectives of instruction, and be compatible with the instructional approaches used.
3. When developing or choosing assessment methods, consideration should be given to the consequences of the decisions to be made in light of the obtained information.
4. More than one assessment method should be used to ensure comprehensive and consistent indications of student performance.
5. Content and language that would generally be viewed as sensitive, sexist, or offensive should be avoided.
6. Assessment instruments translated into a second language or transferred from another context or location should be accompanied by evidence that inferences based on these instruments are valid for the intended purpose.

### 0.2.2 Collecting Assessment Information

1. Students should be told why assessment information is being collected and how this information will be used.
2. An assessment procedure should be used under conditions suitable to its purpose and form.
3. In assessments involving observations, checklists, or rating scales, the number of characteristics to be assessed at one time should be small enough and concretely described so that the observations can be made accurately.
4. The directions provided to students should be clear, complete, and appropriate for the ability, age, and grade level of the students.
5. In assessment involving selection items (e.g., true-false, multiple-choice), the directions should encourage students to answer all items without threat of penalty.
6. When collecting assessment information, interactions with students should be appropriate and consistent.
7. Unanticipated circumstances that interfere with the collection of assessment information should be noted and recorded.
8. A written policy should guide decisions about the use of alternate procedures for collecting assessment information from students with special needs and students whose proficiency in the language of instruction is inadequate for them to respond in the anticipated manner.

### 0.2.3 Judging and Scoring Student Performance

1. Before an assessment method is used, a procedure for scoring should be prepared to guide the process of judging the quality of a performance or product, the appropriateness of an attitude or behavior, or the correctness of an answer.
2. Before an assessment method is used, students should be told how their responses or the information they provide will be judged or scored.
3. Care should be taken to ensure that results are not influenced by factors that are not relevant to the purpose of the assessment.
4. Comments formed as part of scoring should be based on the responses made by the students and presented in a way that students can understand and use them.
5. Any changes made during scoring should be based upon a demonstrated problem with the initial scoring procedure. The modified procedure should then be used to rescore all previously scored responses.
6. An appeal process should be described to students at the beginning of each school year or course of instruction that they may use to appeal a result.

### 0.2.4 Summarizing and Interpreting Results

1. Procedures for summarizing and interpreting results for a reporting period should be guided by a written policy.
2. The way in which summary comments and grades are formulated and interpreted should be explained to students and their parents/guardians.
3. The individual results used and the process followed in deriving summary comments and grades should be described in sufficient detail so that the meaning of a summary comment or grade is clear.
4. Combining disparate kinds of results into a single summary should be done cautiously. To the extent possible, achievement, effort, participation, and other behaviors should be graded separately.
5. Summary comments and grades should be based on more than one assessment result so as to ensure adequate sampling of broadly defined learning outcomes.
6. The results used to produce summary comments and grades should be combined in a way that ensures that each result receives its intended emphasis or weight.
7. The basis for interpretation should be carefully described and justified.
8. Interpretations of assessment results should take account of the backgrounds and learning experiences of the students.
9. Interpretations of assessment results should take account of the backgrounds and learning experiences of the students.
10. Interpretations of assessment results should be made with due regard for limitations in the assessment methods used, problems encountered in collecting the information and judging or scoring it, and limitations in the basis used for interpretation.

### 0.2.5 Reporting Assessment Findings

1. The reporting system for a school or jurisdiction should be guided by a written policy. Elements to consider include such aspects as audiences, medium, format, content, level of detail, frequency, timing, and confidentiality.
2. Written and oral reports should contain a description of the goals and objectives of instruction to which the assessments are referenced.
3. Reports should be complete in their descriptions of strengths and weaknesses of students, so that strengths can be build upon and problem areas addressed.
4. The reporting system should provide for conferences between teachers and parents/guardians. Whenever it is appropriate, students should participate in these conferences.
5. An appeal process should be described to students and their parents/guardians at the beginning of each school year or course of instruction that they may use to appeal a report.
6. Access to assessment information should be governed by a written policy that is consistent with applicable laws and with basic principles of fairness and human rights.
7. Transfer of assessment information from one school to another should be guided by a written policy with stringent provisions to ensure the maintenance of confidentiality.

Source: Joint Advisory Committee. (1993). Principles for Fair Student Assessment Practices for Education in Canada. University Of Alberta. <https://www.wcdsb.ca/wp-content/uploads/sites/36/2017/03/fairstudent.pdf>

## 0.3 Characteristics of sound assessments

Sound assessments 1. arise from and serve clear purposes; 2. arise from and reflect clear and appropriate achievement targets; 3. rely on a proper assessment method, given the purpose and the target; 4. sample student achievement appropriately; and 5. control for all relevant sources of bias and distortion.

Source: Stiggins, R. J. (1995). Assessment Literacy for the 21st Century. *The Phi Delta Kappan, 77*(3), 238–245. JSTOR.

## 0.4 Educational assessment knowledge and skills for teachers

1. Teachers should understand learning in the content area they teach.
2. Teachers should be able to articulate clear learning intentions that are congruent with both the content and depth of thinking implied by standards and curriculum goals, in such a way that they are attainable and assessable.
3. Teachers should have a repertoire of strategies for communicating to students what achievement of a learning intention looks like.
4. Teachers should understand the purposes and uses of the range of available assessment options and be skilled in using them.
5. Teachers should have the skills to analyze classroom questions, test items and performance assessment tasks to ascertain the specific knowledge and thinking skills required for students to do them.
6. Teachers should have the skills to provide effective, useful feedback on student work.
7. Teachers should be able to construct scoring schemes that quantify student performance on classroom assessments into useful information for decisions about students, classrooms, schools, and districts. These decisions should lead to improved student learning, growth, or development.
8. Teachers should be able to administer external assessments and interpret their results for decisions about students, classrooms, schools, and districts.
9. Teachers should be able to articulate their interpretations of assessment results and their reasoning about the educational decisions based on assessment results to the educational populations they serve (student and his/her family, class, school, community).
10. Teachers should be able to help students use assessment information to make sound educational decisions.
11. Teachers should understand and carry out their legal and ethical responsibilities in assessment as they conduct their work.

Source: Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. Educational Measurement: Issues and Practice, 30, 3–12. <https://doi.org/10/cwcqj4>

## 0.5 Classroom Assessment Standards for PreK-12 Teachers

### 0.5.1 Foundation

1. Assessment Purpose: Classroom assessment practices should have a clear purpose that supports teaching and learning.
2. Learning Expectations: Learning expectations should form the foundation for aligning classroom assessment practices with appropriate instruction and learning opportunities for each student.
3. Assessment Design: The types and methods of classroom assessment used should clearly allow students to demonstrate their learning.
4. Student Engagement in Assessment: Students should be meaningfully engaged in the assessment process and use of the assessment evidence to enhance their learning.
5. assessment Preparation: Adequate teacher and student preparation in terms of resources, time, and learning opportunities should be a part of classroom assessment practices,
6. Informed Students and Parents/Guardians: The purposes and uses of classroom assessment should be communicated to students and, when appropriate, parents/guardians.

### 0.5.2 Use

1. Analysis of Student Performance: The methods for analyzing evidence of student learning should be appropriate for the assessment purpose and practice.
2. Effective Feedback: Classroom assessment practices should provide timely and useful feedback to improve student learning.
3. Instructional Follow-up: Analysis of student performance should inform instructional planning and next steps to support ongoing student learning.
4. Grades and Summary Comments: Summative grades and comments should reflect student achievement of the learning expectations.
5. Reporting: Assessment reports should be based on a sufficient body of evidence and provide a summary of a student’s learning in a clear, timely, accurate, and useful manner.

### 0.5.3 Quality

1. Cultural and linguistic diversity: Classroom assessment practices should be responsive to and respectful of the cultural and linguistic diversity of students and their communities.
2. Exceptionality and special education: Classroom assessment practices should be appropriately differentiated to meet the specific educational needs of all students.
3. Unbiased and fair assessment: Classroom assessment practices and subsequent decisions should be free from all factors unrelated to the intended purposes of the assessment.
4. Reliability and validity: Classroom assessment practices should provide consistent, dependable, and appropriate information that supports sound interpretations and decisions about each student’s knowledge and skills,
5. Reflection: Classroom assessment practices should be monitored and revised to improve their overall quality.

Klinger, D., McDivitt, P., Howard, B., Rogers, T., Munoz, M., & Wylie, C. (2015). Classroom Assessment Standards for PreK-12 Teachers. Joint Committee on Standards for Educational Evaluation. <https://www.amazon.ca/Classroom-Assessment-Standards-PreK-12-Teachers-ebook/dp/B00V6C9RVO?asin=B00V6C9RVO&revisionId=d45424dd&format=1&depth=1>

AFT, NCME, & NEA. (1990). *Standards for Teacher Competence in Educational Assessment of Students*.

Barnett, R. (2007). *A Will to Learn: Being a Student in an Age of Uncertainty*. McGraw-Hill Education.

Bernstein, B. (1999). Vertical and Horizontal Discourse: An Essay. *British Journal of Sociology of Education*, *20*(2), 157–173. <https://doi.org/ftmsvc>

Biggs, J. (1999). What the Student Does: Teaching for enhanced learning. *Higher Education Research & Development*, *18*(1), 57–75. <https://doi.org/drgphk>

Biggs, J. B. (1993). From Theory to Practice: A Cognitive Systems Approach. *Higher Education Research & Development*, *12*(1), 73–85. <https://doi.org/ccdmd9>

Birenbaum, M. (2014). Conceptualizing Assessment Culture in School. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing Assessment for Quality Learning* (Vol. 1, pp. 285–302). Springer Netherlands. <https://doi.org/10.1007/978-94-007-5902-2_18>

Birenbaum, M. (2016). Assessment Culture Versus Testing Culture: The Impact on Assessment for Learning. In D. Laveault & L. Allal (Eds.), *Assessment for Learning: Meeting the Challenge of Implementation* (Vol. 4, pp. 275–292). Springer International Publishing. <https://doi.org/10.1007/978-3-319-39211-0_16>

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. <https://doi.org/fpnss4>

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>

Bloom, B. (1968). Learning for Mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Evaluation Comment*, *1*(2), 12.

Broadfoot, P. (2016). Assessment for Twenty-First-Century Learning: The Challenges Ahead. In M. J. Spector, B. B. Lockee, & M. D. Childress (Eds.), *Learning, Design, and Technology* (pp. 1–23). Springer International Publishing. <https://doi.org/10.1007/978-3-319-17727-4_64-1>

Brookhart, S. M. (2011). Educational Assessment Knowledge and Skills for Teachers. *Educational Measurement: Issues and Practice*, *30*, 3–12. <https://doi.org/cwcqj4>

Brown, G. T. L. (2004). Teachers’ conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, *11*(3), 301–318. <https://doi.org/10.1080/0969594042000304609>

Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in confucian-heritage settings*. Routledge.

Committee, J. A. (1993). *Principles for Fair Student Assessment Practices for Education in Canada*. University Of Alberta.

Crooks, T. J. (1988). The Impact of Classroom Evaluation Practices on Students. *Review of Educational Research*, *58*(4), 438–481. <https://doi.org/dvd8nf>

DeLuca, C. (2012). Preparing teachers for the age of accountability: Toward a framework for assessment education. *Action in Teacher Education*, *34*, 576–591. <https://doi.org/10.1080/01626620.2012.730347>

DeLuca, C., Chavez, T., & Cao, C. (2013). Establishing a foundation for valid teacher judgement on student learning: The role of pre-service assessment education. *Assessment in Education: Principles, Policy & Practice*, *20*(1), 107–126. <https://doi.org/gj5v98>

DeLuca, C., Coombs, A., Macgregor, S., & Rasooli, A. (2019). Toward a differential and situated view of assessment literacy: Studying teachers’ responses to classroom assessment scenarios. *Frontiers in Education*, *4*. <https://doi.org/gh5k63>

DeLuca, C., & Klinger, D. (2010). Assessment literacy development: Identifying gaps in teacher candidates’ learning. *Assessment in Education: Principles, Policy & Practice*, *17*, 419–438. <https://doi.org/c2cw5r>

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, *21*, 248–266. <https://doi.org/gfgtsg>

DeLuca, C., Rickey, N., & Coombs, A. (2021). Exploring assessment across cultures: Teachers’ approaches to assessment in the U.S., China, and Canada. *Cogent Education*, *8*(1), 1921903. <https://doi.org/gjxvc7>

DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2016). Teachers’ approaches to classroom assessment: A large-scale survey. *Assessment in Education: Principles, Policy & Practice*, *25*, 355–375. <https://doi.org/gh5k6p>

DeLuca, C., Willis, J., Cowie, B., Harrison, C., Coombs, A., Gibson, A., & Trask, S. (2019). Policies, Programs, and Practices: Exploring the Complex Dynamics of Assessment Education in Teacher Education Across Four Countries. *Frontiers in Education*, *4*, 132. <https://doi.org/gh5k2r>

Earl, L. M. (2013). *Assessment as learning: Using classroom assessment to maximize student learning* (Second edition). Corwin Press.

Eyal, L. (2012). Digital Assessment Literacy the Core Role of the Teacher in a Digital Environment. *Journal of Educational Technology & Society*, *15*(2), 37–49.

Fletcher, R. B., Meyer, L. H., Anderson, H., Johnston, P., & Rees, M. (2012). Faculty and Students Conceptions of Assessment in Higher Education. *Higher Education*, *64*(1), 119–133. <https://doi.org/ctccpq>

Forum, W. E. (2020). *The Future of Jobs Report 2020*. World Economic Forum.

Garrison, D. R., & Akyol, Z. (2015). Thinking Collaboratively in Educational Environments: Shared Metacognition and Co-Regulation in Communities of Inquiry. In J. Lock, P. Redmond, & P. A. Danaher (Eds.), *Educational Developments, Practices and Effectiveness* (pp. 39–52). Palgrave Macmillan UK. <https://doi.org/10.1057/9781137469939_3>

Harlen, W., & Deakin Crick, R. (2002). A systematic review of the impact of summative assessment and tests on students’ motivation for learning. *Research Evidence in Education Library*, *1*, 151.

Hébert, C. (2021). Online Remote Proctoring Software in the Neoliberal Institution: Measurement, Accountability, and Testing Culture. *In Education*, *27*(1), 23–40. <https://doi.org/10.37119/ojs2021.v27i1.507>

James, M. (2008). Assessment and learning. In S. Swaffield (Ed.), *Unlocking Assessment*. David Fulton Publishers. <https://doi.org/10.4324/9780203930939>

Joughin, G. (Ed.). (2009). *Assessment, Learning and Judgement in Higher Education*. Springer Netherlands. <https://doi.org/10.1007/978-1-4020-8905-3>

Klinger, D., McDivitt, P., Howard, B., Rogers, T., Munoz, M., & Wylie, C. (2015). *Classroom Assessment Standards for PreK-12 Teachers*. Joint Committee on Standards for Educational Evaluation.

Lipnevich, A. A., Guskey, T. R., Murano, D. M., & Smith, J. K. (2020). What do grades mean? Variation in grading criteria in American college and university courses. *Assessment in Education: Principles, Policy & Practice*, *27*(5), 480–500. <https://doi.org/ghjw3k>

Looney, A., Cumming, J., van der Kleij, F. M., & Harris, K. (2017). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, *25*, 442–467. <https://doi.org/gfkfk6>

Massey, K. D., DeLuca, C., & LaPointe-McEwan, D. (2020). Assessment Literacy in College Teaching: Empirical Evidence on the Role and Effectiveness of a Faculty Training Course. *To Improve the Academy*, *39*(1). <https://doi.org/gj5ngz>

Mislevy, R. J. (1994). Test theory reconcieved. *ETS Research Report Series*, *1994*(1), i–38. <https://doi.org/gjm236>

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, *14*(2), 149–170. <https://doi.org/10.1080/09695940701478321>

Offerdahl, E. G., & Tomanek, D. (2011). Changes in instructors’ assessment thinking related to experimentation with new strategies. *Assessment and Evaluation in Higher Education*, *36*(7), 781–795. <https://doi.org/d89z7p>

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. National Academies Press. <https://doi.org/10.17226/10019>

Pellegrino, J. W., & Quellmalz, E. S. (2010). Perspectives on the Integration of Technology and Assessment. *Journal of Research on Technology in Education*, *43*(2), 119–134. <https://doi.org/ggfh8z>

Popham, W. J. (2009). Assessment Literacy for Teachers: Faddish or Fundamental? *Theory Into Practice*, *48*(1), 4–11. <https://doi.org/djcn3z>

Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. <https://doi.org/10.1016/j.tics.2016.07.002>

Scriven, M. (1967). [The methodology of evaluation](https://www.ncbi.nlm.nih.gov/pmc/articles/null). In B. O. Smith (Ed.), *Perspectives of curriculum evaluation*. Rand McNally.

Shepard, L. A. (2000). The Role of Assessment in a Learning Culture. *Educational Researcher*, *29*(7), 4–14. <https://doi.org/cw9jwc>

Singh, P. (2002). Pedagogising Knowledge: Bernstein’s theory of the pedagogic device. *British Journal of Sociology of Education*, *23*(4), 571–582. <https://doi.org/10.1080/0142569022000038422>

Skinner, B. (1938). *The behaviour of organisms*. Appleton-Century-Crofts.

Stiggins, R. J. (1991). Assessment Literacy. *The Phi Delta Kappan*, *72*(7), 534–539.

Stiggins, R. J. (1995). Assessment Literacy for the 21st Century. *The Phi Delta Kappan*, *77*(3), 238–245.

Thorndike, E. L. (1905). *The Elements of psychology*. A.G. Seiler.

Timmis, S., Broadfoot, P., Sutherland, R., & Oldfield, A. (2016). Rethinking assessment in a digital age: Opportunities, challenges and risks. *British Educational Research Journal*, *42*(3), 454–476. <https://doi.org/gftz95>

Vygotsky, L. S. (1978). *Mind in society* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.; A. R. Luria, Trans.). Harvard University Press.

Webb, M., & Ifenthaler, D. (2018). Assessment as, for, and of Twenty-First Century Learning Using Information Technology: An Overview. In J. Voogt, G. Knezek, R. Christensen, & K.-W. Lai (Eds.), *Second Handbook of Information Technology in Primary and Secondary Education* (pp. 581–600). Springer International Publishing. <https://doi.org/10.1007/978-3-319-71054-9_37>

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers’ assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, *40*(2), 241–256. <https://doi.org/gh5k7d>