Approaches to Assessment in Higher Education

Colin Madland

2022-03-24

Table of Contents

# Welcome

This site hosts the open development of my PhD dissertation at University of Victoria. It is an active work in progress.

Please contact me at cmadland [at] uvic [dot] ca if you have questions or comments. Alternately, you can annotate with hypothes.is, which is active in the top, right corner. Keep in mind, that with regular updates, your annotation may very well become orphaned.

# Introduction

Assessing learning is both a critical component of the work of teaching in higher education and also a major factor in learners’ experiences of higher education ([J. Biggs, 1999](#ref-biggsWhatStudentDoes1999); [Woldeab & Brothen, 2019](#ref-woldeab21stCenturyAssessment2019)). Faculty and instructors’ approaches to assessment in higher education are shaped by a variety of factors, including the approaches to assessment that they experienced as learners ([Lipnevich et al., 2020](#ref-lipnevichWhatGradesMean2020); [Massey et al., 2020](#ref-masseyAssessmentLiteracyCollege2020)), pressure from our increasingly technological society to integrate digital tools into the teaching and learning process ([Pellegrino & Quellmalz, 2010](#Xd3220b3c4c74762840488f1ab7714e7e895bc2e)), and the changing needs of 21st-century employers and objectives of higher education institutions, who seek employees and graduates with demonstrated ability in collaboration, creative problem-solving, analytical thinking, and the ability to learn ([Forum, 2020](#Xaf0f34faf6db1662e8da0b9850d3b9cfe5dbeed); [Shute et al., 2010](#ref-shute21stCenturyAssessment2010)), competencies not easily measured in traditional testing formats.

Massey et al., ([2020](#ref-masseyAssessmentLiteracyCollege2020)) contend that instructors in higher education typically have few opportunities to engage in formal preparation for the task of assessing learning, and consequently there is high variability in how instructors assess learning in their courses. This is congruent with Coombs et al.’s ([2020](#ref-coombsSeaSeaCanadian2020)) findings that even preservice teachers in teacher preparation programs are often unprepared for the challenge of assessing learning. If pre-service teachers, who complete a program of formal academic preparation for teaching, are under-prepared, it follows that those who exit doctoral programs with no formal preparation for teaching or assessment ([Lipnevich et al., 2020](#ref-lipnevichWhatGradesMean2020)) will be even less adequately prepared. This lack of formal preparation generally means that higher education instructors assess learning in the only way they know how, which is to follow the example of their supervisors and professors from graduate school.

In her important article, Shepard ([2000](#ref-shepardRoleAssessmentLearning2000)) argues that traditional 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) 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)), 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 higher education instructors who resisted changing their assessments to match the new curricular goals ([Shepard, 2000](#ref-shepardRoleAssessmentLearning2000)). 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 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 the continual learning as being priorities for 21st century employers ([Forum, 2020](#Xaf0f34faf6db1662e8da0b9850d3b9cfe5dbeed)). Models of assessment which mimic and re-inscribe traditional assessment practices and 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 higher education ([Broadfoot, 2016](#X51a879b1b801bbc5ad6e7fc588f4456f2973b2d); [Crooks, 1988](#ref-crooksImpactClassroomEvaluation1988); [Pellegrino & Quellmalz, 2010](#Xd3220b3c4c74762840488f1ab7714e7e895bc2e); [Timmis et al., 2016](#X9783eef1c37e492f1192ce9092108eb526b7f44)).

Furthermore, pressures from our increasingly technological society have also impacted instructors’ approaches to their assessment practice. O’Donnell ([2020](#Xb8ef97dcd64a46920c7df0026dc0b714b2ee92d)) argues that *all* learning in higher education is now mediated in some way by technology, even if it is the superficial use of a word processor an instructor uses to create course materials or a learner uses to compose an essay. Even so, he claims that the use of technology often does little more than increase the efficiency (a term which he declines to define) of existing practices (an example might be reducing the time it takes to score a selected-response exam by using bubble sheets for examinee responses). This critique echoes Timmis et al. ([2016](#X9783eef1c37e492f1192ce9092108eb526b7f44)) who argue that prioritizing superficial characteristics of technology, like ‘efficiency,’ comes at the cost of more innovative applications.

Additionally, the datafication of higher education has made very large data sets available to individual instructors as well as to learning technology administrators. This is often in the form of log data from learning management systems (LMS) which has been used to explore relationships between learner behaviours in the LMS and achievement ([Pardo & Reimann, 2020](#ref-pardoBidirectionalEffectData2020); [Stadler et al., 2020](#ref-stadlerFirstEqualsLog2020)). Shute and Rahimi reported on their exploration of what they call “stealth assessments” ([2021, p. 4](#ref-shuteStealthAssessmentCreativity2021)) where large amounts of data are collected as learners interact in game-based learning environments. They argue that stealth assessments, which learners do not notice because they are woven seamlessly into the learning materials and automatically scored based help to alleviate test anxiety for learners, leading to greater achievement ([Shute et al., 2010](#ref-shute21stCenturyAssessment2010)). Milligan writes about the optimism with which early learning analytics researchers advocated for using big data that ([2020](#X176a232588f05c4bc332011296c94f6bd1d0211))

## Problem to be Researched

Given the plurality of factors influencing higher education instructors identified above and the relative paucity of published research on assessment in higher education compared to K-12 ([Lipnevich et al., 2020](#ref-lipnevichWhatGradesMean2020); [Massey et al., 2020](#ref-masseyAssessmentLiteracyCollege2020)), this dissertation will unfold over the course of three independent publications. The common thread running through the three publications will be a focus on the problem of the alignment gap between the assessment practices of many higher education instructors and the goals and objectives of modern higher education institutions. Each publication will explore this topic from different viewpoints, or lenses. The first paper will be a detailed review of the literature related to assessment and the impacts of technology on assessment in higher education using DeLuca et al.’s ([2016a](#Xedee9bc10fe7ddda531b14b05542852ae9f5f1e)) approaches to classroom assessment framework. The second paper will be a quantitative analysis of DeLuca et al.’s *Approaches to Classroom Assessment Inventory* ([ND](#ref-delucaACAIInstrumentSpecificationsND)) revised for a technologically-mediated higher education environment. The third paper will consist of a synthesis of findings from the first two papers with a view to reimagining assessment in 21st-century higher education and consideration given to the ethics of modern trends in technology.

~~The author notes that there are myriad examples of formal and informal initiatives in higher education which have not realized significant local or systemic change (~~[~~Broadfoot, 2016~~](#X51a879b1b801bbc5ad6e7fc588f4456f2973b2d)~~;~~ [~~Earl, 2013~~](#ref-earlAssessmentLearningUsing2013)~~). Instructors are highly resistant to changing their assessment practices, in part, as Broadfoot claims, because assessment is so important, but also, as will be discussed later, because the approaches that individual instructors take to assessing the work of learners are driven by complex forces (~~[~~Black & Wiliam, 1998~~](#ref-blackAssessmentClassroomLearning1998)~~;~~ [~~DeLuca et al., 2019~~](#ref-delucaDifferentialSituatedView2019)~~;~~ [~~Stiggins, 1991~~](#ref-stigginsAssessmentLiteracy1991)~~;~~ [~~Willis et al., 2013~~](#X731be6e2e0895c79e0f034fc829fd623ca4b0d3)~~).~~

## Purpose of the Research

Following previous research by DeLuca and colleagues ([DeLuca et al., 2021](#Xcffd07f338ae9c1e0f3a4638f84c8ddd08fd0be); [DeLuca, Valiquette, et al., 2016](#X773f46dba91e2266ea33fc285a547bc86776461); [DeLuca, LaPointe-McEwan, et al., 2016a](#Xedee9bc10fe7ddda531b14b05542852ae9f5f1e)) in the K-12 sector, and Massey et al. ([2020](#ref-masseyAssessmentLiteracyCollege2020)) in the higher education sector, the purpose of this research is to investigate current assessment literacies and practices among higher education instructors and the impacts of those approaches on learners. In order to better respond to actual assessment practices, it is critical to understand the conceptions of HE instructors with respect to assessment ([DeLuca et al., 2021](#Xcffd07f338ae9c1e0f3a4638f84c8ddd08fd0be); [Offerdahl & Tomanek, 2011](#X568de4367a2adb3cc7734bdfe3ece898df67951)). Similarly, due to the significant influence assessment practices have on learners and learning, understanding the relationship between instructors’ assessment conceptions and practices and the experience of learners will be important in order to provide a foundation for moving into the remainder of the 21st century with assessment practices aligned with both pedagogical models and learner contexts.

## Research Questions

Assessment in HE is a tremendously complex enterprise with myriad forces influencing the approaches instructors take in their assessment practice, which, in turn, becomes one of the myriad forces influencing the approaches learners take in their learning activities. Instructors conceptions of assessment are often derived from their past experiences as learners who were the objects of assessment, leading to an over-reliance on summative assessment strategies which may not obtain sufficient levels of reliability or validity. Increasing assessment literacy among HE instructors is a necessary, but not likely sufficient, step in advancing fair and balanced assessment practices in an increasingly technology-mediated HE landscape. These concepts lead to the following research questions:

1. Are there distinct patterns in higher education instructors’ approaches to assessment in Canada?
2. Does the prevalence of these patterns differ by:
   * instructors’ levels of experience in teaching face-to-face versus online?
   * instructors’ levels of experience using technology?
3. What factors influence instructors’ approaches to assessment?
4. How do different assessment strategies affect learners’ experiences?

## Significance of the Research

# A Review of the Literature on Assessment in Higher Education

## Structure of the Paper

This paper will begin with establishing a clear definition of assessment followed by an analysis of the literature related to the concept of *assessment literacy* (AL) ([DeLuca et al., 2019](#ref-delucaDifferentialSituatedView2019)) as it relates to the approaches to assessment ([DeLuca, LaPointe-McEwan, et al., 2016a](#Xedee9bc10fe7ddda531b14b05542852ae9f5f1e)) taken by HE instructors and the perceptions and experiences of HE learners. The last section will introduce a framework for understanding assessment in HE, situated within Biggs’ ([1999](#ref-biggsWhatStudentDoes1999); [1993](#ref-biggsTheoryPracticeCognitive1993)) 3P model of teaching and learning, and research questions which emerge from the literature.

## Defining Assessment

There are deep and rich bodies of literature addressing educational assessment *writ large*, both from a summative, psychometric perspective, and from a formative perspective. There is notably more research in the K-12 context, especially in relation to teacher preparation, compared to HE.

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 thorough review of the literature, 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 has remained high [Lipnevich et al.](#ref-lipnevichWhatGradesMean2020) ([2020](#ref-lipnevichWhatGradesMean2020)).

The National Research Council’s (NRC) 2001 report *Knowing what students know,* advanced understanding of assessment with their definition of assessment as “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)) or, more simply, “reasoning from evidence” ([Pellegrino et al., 2001, p. 43](#ref-pellegrinoKnowingWhatStudents2001)), based on Mislevy’s assertion that “test theory is machinery for reasoning from students’ behavior to conjectures about their competence, as framed in a particular conception of competence.” ([1994, p. 4](#ref-mislevyTestTheoryReconcieved1994)). 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. Hence, *all* summative assessments are probablistic, not deterministic.

More recently, Earl ([2013](#ref-earlAssessmentLearningUsing2013)) further clarified the role assessment can play in learning by highlighting a distinction between assessment *of* learning (summative assessment) and assessment *for* learning (formative assessment by way of feedback) and also distinguishing both of those from assessment *as* learning (a subset of assessment *for* learning in which learners employ metacognitive skills to regulate their own learning tasks). Earl’s delineation between types of assessment reflects the modern view that assessment and learning are, or ought to be, tightly integrated. This relationship will be explored in more detail in relation to Biggs’ 3P Model of Teaching and Learning ([1999](#ref-biggsWhatStudentDoes1999), [1996](#X6d502c1fbcba67bdb61d23bc33194bf5c0f1ecb)).

The definitions of assessment above are typically understood as being *classroom assessment*, language which is more readily applied to face-to-face K-12 learning environments as opposed to HE environments mediated by technology. For the purposes of this paper, I will consider assessment *of* learning and summative assessment to be essentially synonymous, and I will differentiate between assessment *for* and *as* learning. I will use the term *classroom assessment* to differentiate from *large-scale assessment*, understood to be assessments deployed at levels above individual classrooms, such as school-, system-, or provincial/federal-levels, and I will use *online assessment* or *technology-mediated assessment* to refer specifically to classroom assessment in learning environments mediated by technology whether the learners are remote or not.

## Conceptions of Assessment

Instructors in HE typically receive little formal preparation in either teaching practices or assessment during their graduate studies ([Lipnevich et al., 2020](#ref-lipnevichWhatGradesMean2020); [Massey et al., 2020](#ref-masseyAssessmentLiteracyCollege2020)). Consequently, their own practice tends to follow from what they experienced as learners, which likely emphasized high-stakes summative tests which were either in alignment with outdated pedagogical practices or out of alignment with modern pedagogical practices. These prior conceptions of assessment carry significant weight in how HE instructors approach the assessment of learners in their own courses.

Instructors bring to an educational environment a host of influences related to their institutional context, their past experiences with assessment, their own course policies, and affective beliefs or conceptions about assessment and its purposes ([J. B. Biggs, 1993](#ref-biggsTheoryPracticeCognitive1993); [G. T. L. Brown et al., 2011](#X7545160f175b9eeeaf92f7a1f4f6c5a1f4f7d20)). These influences play a significant role in determining the approaches taken by both instructors with respect to assessment and learners with respect to how they approach learning tasks ([G. T. L. Brown et al., 2011](#X7545160f175b9eeeaf92f7a1f4f6c5a1f4f7d20)). DeLuca et al. ([2013](#X32b95a94fb3076f8d0c577bd9276160db2e95e5)) argue that there are categories of conceptions exhibited by K-12 preservice teachers: assessment as testing, assessment as format, assessment purpose, and assessment as process. These conceptions are seen as increasingly complex, with those who see assessment as testing believing that assessment is primarily concerned with summative assessment of learning, usually using teacher-created selected-response tests. Those who see assessment as format tend to focus on whether the assessment is a “performance, product, or objectively-scored assessment” (p. 110). Assessment as purpose is delineated according to the summative/formative binary or Earl’s ([2013](#ref-earlAssessmentLearningUsing2013)) assessment *of/for/as* learning model. DeLuca et al. also identify other purposes of assessment such as accountability, gatekeeping, and teacher evaluation. Lastly, assessment as process, which is based on the National Research Council’s description of assessment being a process of reasoning from evidence ([2001](#ref-pellegrinoKnowingWhatStudents2001)).

Fletcher et al. ([2012](#Xa55a5d631b402fa2c3fda6e3997bc5d48ccd027)) used Brown’s ([2017](#X398061bedf4ca6ea011494eb395f9026963cd42)) abridged *Conceptions of Assessment* (CoA) questionnaire to measure learners’ and instructors’ conceptions as follows: “assessment makes institutions accountable, assessment makes students accountable, assessment describes improvements in student abilities, assessment improves student learning, assessment improves teaching, assessment is valid, assessment is irrelevant and bad, assessment is irrelevant and ignored, and assessment is irrelevant and inaccurate” (p. 122). They report that instructors were more likely than learners to view assessment as consistent and trustworthy methods to understand and improve learning and that learners were more likely to have negative views of assessment and viewed it as a measure of student and institutional accountability.

Massey et al. ([Massey et al., 2020](#ref-masseyAssessmentLiteracyCollege2020)) used DeLuca et al.’s ([2013](#X32b95a94fb3076f8d0c577bd9276160db2e95e5)) framework of conceptions in their study of HE instructors’ conceptions of assessment before and after an instructional development course focussed on assessment. They also considered the idea that there are two general orientations towards assessment in HE, an “assessment culture” and a “testing culture” [Massey et al.](#ref-masseyAssessmentLiteracyCollege2020) ([2020](#ref-masseyAssessmentLiteracyCollege2020)). They report that they saw significant shifts in participants’ conceptions of assessment from more simplistic views of assessment as testing pre-treatment, to more complex and nuanced views or assessment as process post-treatment.

From the literature, it is clear that instructors’ conceptions of assessment are deeply influenced by many internal and external factors and, especially in HE, where there are few constraints on assessment practice ([Lipnevich et al., 2020](#ref-lipnevichWhatGradesMean2020)), there are many ways to describe or delineate different conceptions of assessment. Accordingly, there are multiple ways to conceptualize the skills and dispositions that comprise the idea of *assessment literacy*.

## Assessment Literacy

The idea of AL is relatively recent in the K-12 literature and is nascent and under-theorized with respect to HE contexts ([Medland, 2015](#X5c86252aa4ff917d9cc0a272767d26b88512cd6)). AL has been defined variously as “the skills and knowledge teachers require to measure and support student learning through assessment” ([DeLuca, LaPointe-McEwan, et al., 2016a](#Xedee9bc10fe7ddda531b14b05542852ae9f5f1e)), “a basic understanding of educational assessment and related skills to apply such knowledge to various measures of student achievement” ([Xu & Brown, 2016](#ref-xuTeacherAssessmentLiteracy2016)), “an individual’s understandings of the fundamental assessment concepts and procedures deemed likely to influence educational decisions” ([Popham, 2011](#X2ef0bb9d5e3fccea0cf8ddd6c54b6184255abbf)) and “a dynamic context-dependent social practice that involves teachers articulating and negotiating classroom and cultural knowledges with one another and with learners, in the initiation, development and practice of assessment to achieve the learning goals of students” ([Willis et al., 2013](#X731be6e2e0895c79e0f034fc829fd623ca4b0d3)). Key to these definitions are the ideas that AL is a complex, multi-faceted construct, that AL requires *adequate* (not high) levels of psychometric or statistical analyses, and that it is intended to enable learner success.

The recognition of AL 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 AL have tended to be based on sets of standards to which K-12 teachers are obligated. The first set of standards was the *Standards for Teacher Competence in Educational Assessment of Students* (the *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 ([AFT et al., 1990](#ref-StandardsTeacherCompetence1990)). The *Standards* are a list of seven skills expected of teachers:

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.

Shortly after the publication of the *Standards*, the term *assessment literacy* appeared in the literature with Stiggins’ ([1991](#ref-stigginsAssessmentLiteracy1991)) article called *Assessment Literacy*. Stiggins initial article was an account of his observation that teacher education programs at the time spent very little time training teachers in the methods and dispositions of educational measurement. Stiggins followed this with another article ([Stiggins, 1995](#ref-stigginsAssessmentLiteracy21st1995)) where he outlined five characteristics of sound assessments, which:

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. ([1995, p. 240](#ref-stigginsAssessmentLiteracy21st1995))

At around the same time, 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.

While these three sets of recommendations varied widely in their granularity, all tended to reflect an emphasis on the 20th century conceptions of curriculum which prioritized linear and sequential teaching of knowledge followed sometime later by selected-response tests of knowledge. This required teachers to be literate in the psychometric skills required to administer and interpret these tests ([DeLuca, LaPointe-McEwan, et al., 2016b](#ref-delucaTeacherAssessmentLiteracy2016); [Shepard, 2000](#ref-shepardRoleAssessmentLearning2000); [Xu & Brown, 2016](#ref-xuTeacherAssessmentLiteracy2016)) or assessment *of* learning. One notable distinction between the sets of standards is that the Canadian committee specifically noted their principles could be applied to K-12 as well as higher education, although the latter context would require changes in how assessment data are reported.

Twenty years following the publication of the *Standards*, Brookhart ([2011](#X36e21314cdda5af289172df8b2fdcbf83ba9ea4)) argued that the *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.

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. Despite the similarities to the *Principles for Fair Student Assessment Practices for Education in Canada*, including at least one common committee member, the JCSEE Standards are specifically not intended for use in HE.

As traditional conceptions of assessment and the standards expected of teachers, grounded in behaviourism and the need for objectivity tended to focus on assessment as a set of skill-based competencies to be employed by instructors, so AL could be defined as a set of sequential tasks in which instructors should engage to ensure objectivity and fairness (e.g. ([Natriello, 1987](#X949dc1c922451eed515cca009b0e2211dbab978))). Recently, as curriculum and pedagogy have changed, several researchers have proposed models related to AL grounded in socio-constructivist views of learning ([DeLuca, 2012](#ref-delucaPreparingTeachersAge2012); [Pastore & Andrade, 2019](#ref-pastoreTeacherAssessmentLiteracy2019); [Xu & Brown, 2016](#ref-xuTeacherAssessmentLiteracy2016)). DeLuca’s ([2012](#ref-delucaPreparingTeachersAge2012)) model, developed in the context of the *No Child Left Behind* accountability mandate in K-12 schools in the USA, is a coherent lens through which to understand how a pre-service teacher could develop assessment expertise throughout their teacher education program. DeLuca frames AL within Fostaty Young and Wilson’s ([**fostatyyoungAssessmentLearningICE2000?**](#Xee32f1944353189dca0312300a052f6c74a91f0)) ICE model which presents an integrated progression of learning through three levels of complexity, ideas, connections, and extensions. Underlying the three levels is a foundational level, where teacher candidates come to know and understand assessment as being situated within a model of teaching and learning, experiential and inclusive pedagogies, and an orientation towards continual professional learning. At the *ideas* level, teacher candidates gain expertise in the big ideas related to assessment (theories of learning, planning educational experiences, classroom assessment, and issues of diversity and inclusion). At the *connections* level, teacher candidates begin to construct their own cognitive conceptions of assessment as they integrate their own past experiences with their new knowledge and add their experience in practica. At the *extensions* level, teacher candidates begin to hone their practice as they deepen their understanding of assessment.

Willis et al. ([2013](#X731be6e2e0895c79e0f034fc829fd623ca4b0d3)) describe assessment literacy in alignment with Bernstien’s ([1999](#X4d78eac5dad262ef2a233c8ad9b66b0f5b3cd1c)) idea that there are “horizontal” and “vertical” discourses (p. 159) with respect to assessment. A horizontal discourse is the local, contextualized discourse around assessment which influences local practice, while a vertical discourse is the formalized, structured discourse on assessment in the literature and other more authoritative venues. They argue that any given instructor’s conceptions and practice of assessment will be informed by and negotiated within multiple horizontal and vertical discourses. Similarly, Xu and Brown’s ([2016](#ref-xuTeacherAssessmentLiteracy2016)) model, teacher AL in practise (TALiP) presents a pathway for pre-service teachers to gain expertise in assessment. Based on a synthesis of 100 peer-reviewed publications between 1985 and 2015, they present a six-component model:

1. the knowledge base
2. teacher conceptions of assessment,
3. institutional and socio-cultural contexts,
4. teacher assessment literacy in practice,
5. teacher learning, and
6. teacher identity as assessor. Finally, Pastore and Andrade ([2019](#ref-pastoreTeacherAssessmentLiteracy2019)) developed their model through a Delphi inquiry of 35 international experts in educational assessment and teacher education. They propose a model with three dimensions, conceptual, praxeological, and socio-emotional.

The authors of each of these three models recognize that AL is conceptualized as a multi-dimensional construct encompassing psychometric skills, affective beliefs and values, external and regulatory environments, and socially negotiated practices. However, these models are largely specific to the K-12 environment in general, and more specifically related to the preparation of K-12 teachers. Given that there are very few parallels between the preparation of K-12 teachers for their role and the preparation of HE instructors for theirs, it seems that a framework for understanding AL among HE instructors should consider their general lack of formal preparation for teaching or assessing learning ([Lipnevich et al., 2020](#ref-lipnevichWhatGradesMean2020); [Massey et al., 2020](#ref-masseyAssessmentLiteracyCollege2020)).

Bearman et al. ([2016](#ref-bearmanSupportAssessmentPractice2016)) proposed a model for assessment decision-making in Australian HE, the Assessment Design Decisions Framwork (ADDF) in which the authors acknowledge the difficulty in translating idealized beliefs about assessment into actual practice as well as the lack of literature regarding *how* HE instructors go about designing assessments. While not specifically an AL framework, there are overlaps in terms of the dimensions they identified. They argue, like those previously mentioned, that assessment is a complex and messy process. Unique to their model is the idea that there is often a disconnect between what instructors know or believe to be true about assessment, and how instructors’ practice is impacted (or not) by their beliefs. Additionally, they argue, in alignment with Price et al. ([2011](#ref-priceIfWasGoing2011)), that AL ought to be considered from both instructor and learner perspectives. Their framework is comprised of six dimensions:

1. purposes of assessments
2. contexts of assessments
3. learner outcomes
4. tasks
5. feedback processes
6. interactions

Herppich et al. ([2018](#Xe3aa0c16cb2acc5791f9f0bae56342b4d10012d)) frame a model of assessment competence beginning with the idea that the purpose of educational assessment is to inform both formative (ongoing learning) and summative (credentialling or certification) decisions locating the educational decision subsequent and subordinate to the judgement. They use an example where a teacher observes a learner struggling with a test and might come to an appropriate judgement of a learner’s knowledge, but make an inappropriate instructional decision based on that information. This distinction is useful when, like Herppich and colleagues, the construct under investigation is *assessment competence*, which is observed *after* instructional activities. However, in the present paper, *conceptions of assessment* and *AL* are the relevant constructs and they are embedded in a model of teaching and learning where conceptions of assessment *precede* instructional activities.

DeLuca et al.’s model ([2016a](#Xedee9bc10fe7ddda531b14b05542852ae9f5f1e)), *approaches to classroom assessment* defines and conceptualizes the *assessment* component of Biggs’ 3P model ([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 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.

1. Assessment process

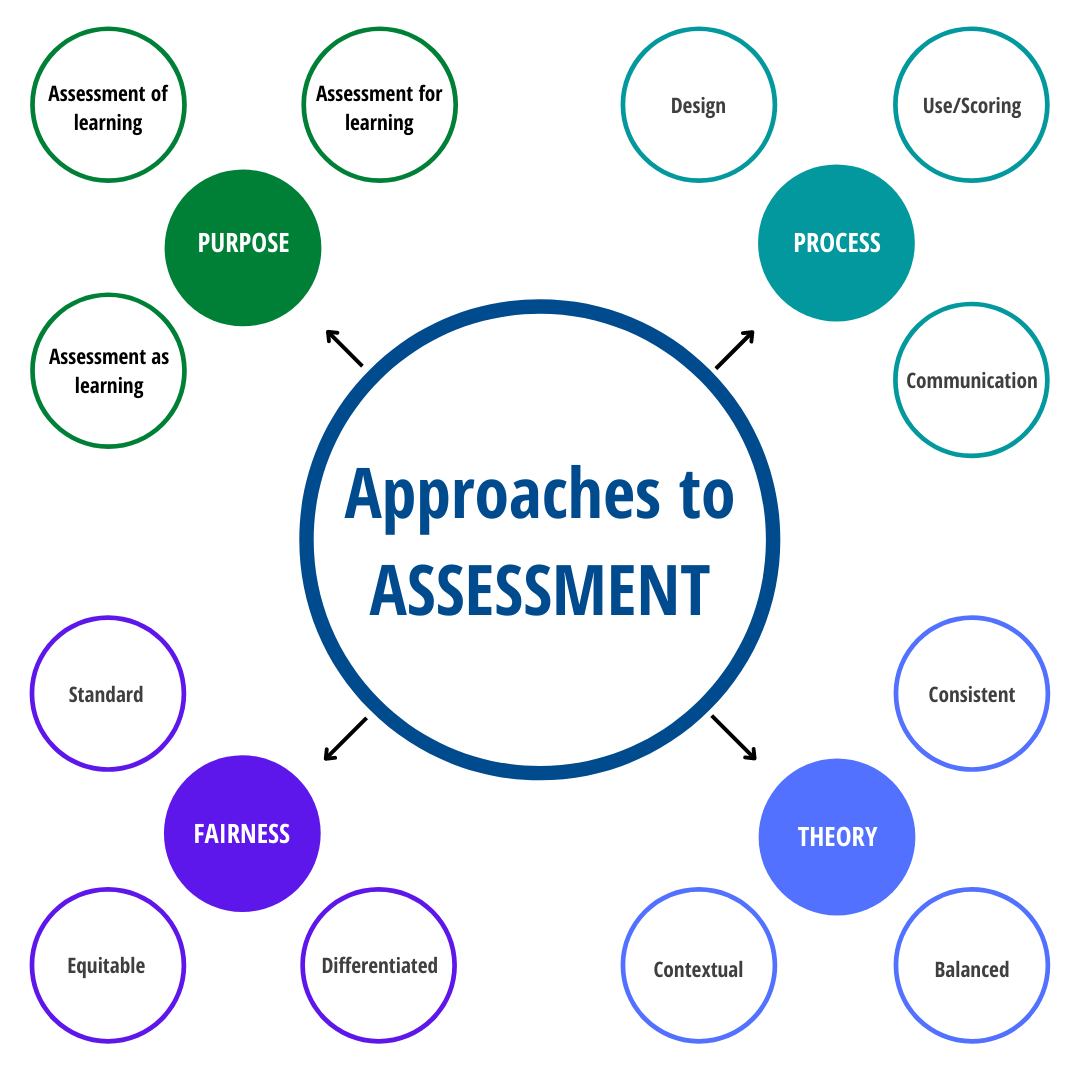
* Design
  + Focuses on the development of reliable assessments and items that measure student learning in relation to learning objectives.
* Use/scoring
  + Focuses of the adjustment and use of scoring protocols and grading schemes to respond to assessment scenarios.
* Communication
  + focuses on the interpretation of assessment results and feedback through communication to students and parents.

1. 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.

1. 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 2. Approaches to Classroom Assessment. ([DeLuca et al., 2021, p. 10](#Xcffd07f338ae9c1e0f3a4638f84c8ddd08fd0be))

## Assessment and Technology

Educational technologies are often viewed and reported on with a distinct positivity bias ([Irvine, 2020](#ref-irvineLandscapeMergingModalities2020)) wherein ‘new’ technologies are presumed to represent progress and will inevitably have a positive effect on learning. This can be seen in the titles given to some initiatives, such as “Technology-Enhanced Assessment” ([Oldfield et al., 2012](#ref-oldfieldAssessmentDigitalAge2012); [Timmis et al., 2016](#X9783eef1c37e492f1192ce9092108eb526b7f44)), “IT-enabled assessment” ([Webb & Ifenthaler, 2018](#ref-webbAssessmentTwentyFirstCentury2018)), or “technology-rich” ([Lin et al., 2020](#Xcf88c9b5dbcbac8629742c94c8d713b40740bc0)). As such, I will use the more neutral term “technology-mediated” to indicate that adding digital technology to an assessment environment does not necessarily improve that environment. Similar to assessment practices being grounded in (both philosophically, as in ‘based upon,’ and figuratively, as in ‘stuck in’) behaviourist conceptions of pedagogy leading to practices that rely heavily on summative approaches to assessment, so too, many educational technologies are grounded in (based upon and stuck in) behaviourist conceptions of pedagogy leading to practices that rely heavily on summative approaches to assessment. This can be seen in the progressively more advanced technologies beginning with Pressey’s teaching machines ([Benjamin, 1988](#ref-benjaminHistoryTeachingMachines1988); [Pressey, 1927](#ref-presseyMachineAutomaticTeaching1927); [A. Watters, 2021](#ref-wattersTeachingMachinesHistory2021)) which was built to automate the process of “drilling” learners in an effort to teach them some concept. The machine needed to be pre-programmed with a series of selected-response questions along with distractors and correct answers. As a learner answered each question, the machine was programmed to match the response to the programmed correct response, and if it matched, the learner was determined to have “mastered” that question and it was dropped from the bank of questions the learner had not yet mastered. If it did not match, the question was cycled back into the bank to be repeated. Clearly, this technology was promoted as a tool to be used to modernize and increase the efficiency of tasks that aligned with the dominant pedagigical paradigm at the time. A second example, although not one marketed directly to schools, but to parents, was the *Speak & Spell*, released in 1978 by Texas Instruments ([Braguinski, 2018](#X37eab5d4961ef5dec4a83729881ef2d687f04ae); [Frantz, 2014](#ref-frantzSpeakSpell2014)), which represented an advance in technology and an increase in efficiency, as the *Speak & Spell* could be programmed to store and reproduce voice recordings of words as well as multiple recordings of feedback messages ([Frantz, 2014](#ref-frantzSpeakSpell2014)). While the *Speak & Spell* was a leap forward in processing power, memory storage, and therefore complexity, the underlying pedagogy remained identical to that of Pressey’s teaching machine ([Audrey Watters, 2015](#ref-wattersSpeakSpellHistory2015)). Moving forward again, and modern technologies are vastly more powerful than teaching machines or the *Speak & Spell* and power very complex adaptive tests, such as the NCLEX-RN, the national licensing exam for Registered Nurses in Canada and the USA ([Smith Glasgow et al., 2019](#X88b1c0c95c5ecb040b63ac0eab37de36c7ef351)). These advances in both hardware and software allow for still greater efficiencies in testing, yet the NCLEX-RN must still be programmed with selected-response questions, their distractors, and correct responses, all still in alignment with behavioural models of pedagogy. The NCLEX-RN is an example of a large-scale, standardized assessment (LSA), so is not parallel to the classroom assessment practices which are the subject of this paper, but I mention it here to draw the distinction between professionally-created LSAs and most instructor-created classroom assessments. The NCLEX-RN is continually revised and updated and reflects very robust psychometric properties (validity and reliability) ([Smith Glasgow et al., 2019](#X88b1c0c95c5ecb040b63ac0eab37de36c7ef351)). More importantly, however, the NCLEX-RN has been updated to “shift away from a primary focus on content and the indirect testing of clinical judgment to a major focus on clinical judgment” ([Caputi, 2019, p. 2](#ref-caputiReflectionsNextGeneration2019)). Caputi, in her reflections on the next-generation NCLEX-RN (NGN) asks 2 questions (the second of which is most relevant here): “1) Are students ready for this type of NCLEX? 2) If our students already pass the NCLEX, can we keep doing the same type of preparation for the NGN?” (p. 2). Her answer to both questions is “No.” She goes on to argue,

So, what can faculty do? I propose that nurse faculty, at all levels of nursing education, revise their curricula to teach a detailed thinking process that students must employ over and over throughout the nursing curriculum. Just as students practice psychomotor skills until they are perfected, they must do so with thinking skills and strategies A new model for teaching clinical judgment is needed", (p. 2, emphasis in original).

In this case, the technology-mediated assessment instrument has been designed to measure 21st century skills, and Caputi recognizes that if pre-service nurses are going to have to pass the NGN, nursing instructors will need to realign their pedagogy. This example illustrates how LSAs exert pressure on instructors and schools to adjust their pedagogy, in the case of nursing, to encourage 21st century pedagogy that teaches thinking skills. The opposite is also true, however, in that LSAs which emphasize the lower-level cognitive skills prized by the behaviourists cause instructors to match their pedagogy to that model ([Caputi, 2019](#ref-caputiReflectionsNextGeneration2019); [Clarke-Midura & Dede, 2010](#X8d225d8bf3a6cb9ee5ffa78c73dfa381c6a7c30); [DeLuca et al., 2021](#Xcffd07f338ae9c1e0f3a4638f84c8ddd08fd0be); [Pellegrino & Quellmalz, 2010](#Xd3220b3c4c74762840488f1ab7714e7e895bc2e)). Further, the American Educational Research Association (AERA), the National Council on Measurement in Education (NCME), and the American Psychological Association (APA) argue that LSAs tend to have other negative effects on education systems, namely the narrowing of curricula (teaching to the test), reduced instructional strategies (previously mentioned), higher dropout rates, and the enactment of policies and practices that increase test scores without increasing learning ([2014](#Xbae586cd6431fa50cc5576da41c7de8abe7eab2)). The NCLEX-RN notwithstanding, many implementations of technology in assessment remain focused on increasing the efficiencies of summative test administration ([Broadfoot, 2016](#X51a879b1b801bbc5ad6e7fc588f4456f2973b2d); [Pellegrino & Quellmalz, 2010](#Xd3220b3c4c74762840488f1ab7714e7e895bc2e); [Webb & Ifenthaler, 2018](#ref-webbAssessmentTwentyFirstCentury2018)).

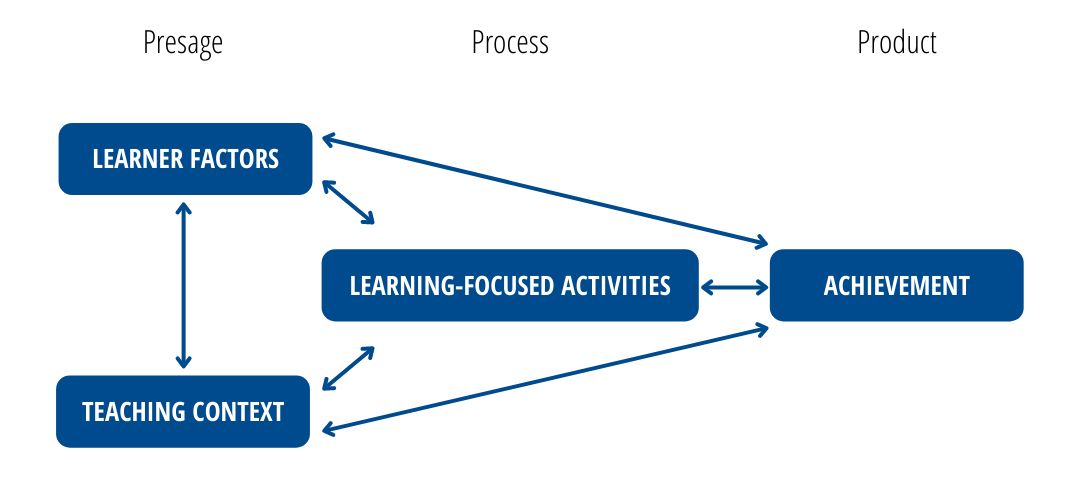
## Impact on Learners

Aside from the effects on learners’ approaches to their learning, assessment, particularly summative assessment, has profound effects in other ways, such as determining status or progression in a course or program, determining eligibility for scholarships and awards, determining career paths and on other affective constructs such as motivation ([Crooks, 1988](#ref-crooksImpactClassroomEvaluation1988)) and anxiety ([Menucha Birenbaum, 2007](#Xed666ae61dbae1b52e6dfe20fa7c7ad7b3c36ec); [Harlen & Deakin Crick, 2002](#ref-harlenSystematicReviewImpact2002)). Jones et al. ([2021](#ref-jonesStudentWellbeingAssessment2021)) report that assessment practices impact learners’ well-being, which they define as including physical and mental health as well as the ability for learners to “fully exercise their cognitive, emotional, physical and social powers, leading to flourishing” (p. 439). They note that summative assessment practices are associated in the literature with “anxiety, depression, disordered eating, self-harm, panic attacks, burnout, … thoughts of suicide … disordered sleep, loss of appetite, physical inactivity, poor physical health, … substance misuse … poorer productivity, motivation and test scores” and that changing from norm-referenced scoring (where learners are ranked relative to each other) to criterion-referenced or pass-fail scoring is associated with lower levels of stress and anxiety among medical students in the USA.

## Theoretical Framework Overview

Given the wide variety of models, mostly in K-12 contexts but increasingly in HE, and the degree to which there appears to be overlap between those models, I take the position that introducing a new model into an already crowded landscape would serve little purpose. Furthermore, models do exist which have been validated empirically ([DeLuca, LaPointe-McEwan, et al., 2016a](#Xedee9bc10fe7ddda531b14b05542852ae9f5f1e)) and the authors of which have indicated that, despite the model being developed for K-12 contexts, adjusting the model to fit the HE context would be a worthwhile contribution to the literature (personal communication).

Here, I present an argument that DeLuca et al.’s idea of *approaches to classroom assessment* ([2016a](#Xedee9bc10fe7ddda531b14b05542852ae9f5f1e)) aligns well with the landscape of teaching, learning, and assessment in HE and may provide a blueprint for assessment reform in HE. In order to situate *Approaches to Classroom Assessment* within the context of teaching and learning in HE, I begin with Biggs’ ([J. Biggs, 1999](#ref-biggsWhatStudentDoes1999); [J. B. Biggs, 1993](#ref-biggsTheoryPracticeCognitive1993)) 3P model of teaching and learning (see Figure 3) which presents three stages of the teaching and learning process. In the first stage, *presage*, Biggs models both learners and the teaching context. The learners’ contexts include their previous learning, cognitive abilities, and their beliefs about the learning context, and all of these influence their preferred approaches to learning. The teaching context includes the instructors beliefs and conceptions of the curriculum and assessment, and also institutional and sometimes regulatory requirements, and each of these influence the instructor’s approach to teaching, and more to the point for this paper, their approach to assessment. The *process* stage represents the activities in which learners engage in order to meet the learning outcomes. Biggs notes that learners tend to either take *surface* approaches, where they use low-level cognitive skills (memorization of facts) in activities that require high-level cognitive skills (analysis or critique), or they take deep approaches, where they use high-level cognitive skills for activities which require them. Learner activity in this stage is influenced by their own backgrounds, abilities, and affective views of the purpose of the task and it is influenced by the choices the instructor makes based on their presage influences. In this way, Earl’s ([2013](#ref-earlAssessmentLearningUsing2013)) delineation between assessment *of*, *for*, and *as* learning can be seen to influence student learning approaches. For example, researchers note that when instructors’ approach prioritizes assessment *of* learning, learners tend to take a surface approach to their learning ([Menucha Birenbaum, 2007](#Xed666ae61dbae1b52e6dfe20fa7c7ad7b3c36ec)). Finally, the *product* stage represents the learners’ achievement in relation to the outcomes. Moving left to right, each of the first two stages represents an input into the system, and the third, an output. However, as the arrows between the components are two-way, the product stage also becomes an input, the results of which are fed back into the system (right to left) informing learners of their status in relation to the outcomes as well as the utility of their learning activities. It also informs the instructor about the nature of the learners’ achievement and the utility of the learning tasks. The *achievement* component provides formative feedback directly to both instructors and learners, informing both about next steps in the learning process. It also provides insight into the learning activities and, when that is fed back to learners, provides metacognitive cues as to the strategies learners might use in future tasks. It also provides information to instructors about the nature of the learning activities and how their approaches to assessment are impacting learning.

 Figure 3. 3P Model of Teaching and Learning. ([J. Biggs, 1999](#ref-biggsWhatStudentDoes1999); [J. B. Biggs, 1993](#ref-biggsTheoryPracticeCognitive1993))

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

Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M., & Molloy, E. (2016). Support for assessment practice: Developing the Assessment Design Decisions Framework. *Teaching in Higher Education*, *21*(5), 545–556. <https://doi.org/gjsn3g>

Benjamin, L. T. (1988). A history of teaching machines. *American Psychologist*, *43*(9), 703–712. <https://doi.org/djgzjr>

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. (1996). Enhancing teaching through constructive alignment. *Higher Education*, *32*(3), 347–364. <https://doi.org/chx3gp>

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. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum & F. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (Vol. 42, pp. 3–29). Kluwer Academic/Plenum Publishers.

Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Opti- mising new modes of assessment: In search of qualities and standards* (pp. 13–36). Kluwer Academic Publishers.

Birenbaum, Menucha. (2007). Assessment and Instruction Preferences and Their Relationship with Test Anxiety and Learning Strategies. *Higher Education*, *53*(6), 749–768. <https://doi.org/cqgvs8>

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

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.

Braguinski, N. (2018). An (An)Archive of Communication: Interactive Toys as Interlocutor. *Communication +1*, *7*(1), 19.

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. (2017). *Teachers Conceptions of Assessment - Secondary Schools Long and Abridged*. <https://doi.org/gj4tz6>

Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers’ conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education*, *27*(1), 210–220. <https://doi.org/c3k8f5>

Caputi, L. J. (2019). Reflections on the Next Generation NCLEX with Implications for Nursing Programs. *Nursing Education Perspectives*, *40*(1), 2–3. <https://doi.org/d8wq>

Clarke-Midura, J., & Dede, C. (2010). Assessment, Technology, and Change. *Journal of Research on Technology in Education*, *42*(3), 309–328. <https://doi.org/ghdnhh>

Coombs, A., Ge, J., & DeLuca, C. (2020). From sea to sea: The Canadian landscape of assessment education. *Educational Research*, 1–17. <https://doi.org/gh5k4z>

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. (ND). *ACAI Instrument, Specifications, and Related Material*. Classroom Assessment Research Team (CART).

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

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016b). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, *28*, 251–272. <https://doi.org/f828mh>

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>

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

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.

Frantz, G. (2014). *The Speak N Spell*. OpenStax CNX.

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.

Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser, J., & Südkamp, A. (2018). Teachers’ assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, *76*, 181–193. <https://doi.org/gfjsvn>

Irvine, V. (2020). The Landscape of merging modalities. *EDUCAUSE Review*, *2020*(4), 40–58.

Jones, E., Priestley, M., Brewster, L., Wilbraham, S. J., Hughes, G., & Spanner, L. (2021). Student wellbeing and assessment in higher education: The balancing act. *Assessment & Evaluation in Higher Education*, *46*(3), 438–450. <https://doi.org/gk36pd>

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.

Lin, Q., Yin, Y., Tang, X., Hadad, R., & Zhai, X. (2020). Assessing learning in technology-rich maker activities: A systematic review of empirical research. *Computers & Education*, *157*, 103944. <https://doi.org/ghgzp9>

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>

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>

Medland, E. (2015). Examining the assessment literacy of external examiners. *London Review of Education*. <https://doi.org/gk5sph>

Milligan, S. (2020). Standards for Developing Assessments of Learning Using Process Data. In M. Bearman, P. Dawson, R. Ajjawi, J. Tai, & D. Boud (Eds.), *Re-imagining University Assessment in a Digital World* (Vol. 7, pp. 179–192). Springer International Publishing. <https://doi.org/10.1007/978-3-030-41956-1_13>

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

Natriello, G. (1987). The Impact of Evaluation Processes on Students. *Educational Psychologist*, *22*(2), 155–175. <https://doi.org/cgqtqx>

O’Donnell, M. (2020). Assessment as and of Digital Practice: Building Productive Digital Literacies. In M. Bearman, P. Dawson, R. Ajjawi, J. Tai, & D. Boud (Eds.), *Re-imagining University Assessment in a Digital World* (Vol. 7, pp. 111–125). Springer International Publishing. <https://doi.org/10.1007/978-3-030-41956-1_9>

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>

Oldfield, A., Broadfoot, P., Sutherland, R., & Timmis, S. (2012). *Assessment in a Digital Age: A research review*. Graduate School of Education, University of Bristol.

Pardo, A., & Reimann, P. (2020). The Bi-directional Effect Between Data and Assessments in the Digital Age. In M. Bearman, P. Dawson, R. Ajjawi, J. Tai, & D. Boud (Eds.), *Re-imagining University Assessment in a Digital World* (Vol. 7, pp. 165–178). Springer International Publishing. <https://doi.org/10.1007/978-3-030-41956-1_12>

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, *84*, 128–138. <https://doi.org/gh5k7b>

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. (2011). Assessment Literacy Overlooked: A Teacher Educator’s Confession. *The Teacher Educator*, *46*(4), 265–273. <https://doi.org/dctz5h>

Pressey, S. L. (1927). A machine for automatic teaching of drill material. *School & Society*, *25*, 549–552.

Price, M., Carroll, J., O’Donovan, B., & Rust, C. (2011). If I was going there I wouldn’t start from here: A critical commentary on current assessment practice. *Assessment & Evaluation in Higher Education*, *36*(4), 479–492. <https://doi.org/d4sz5m>

Scriven, M. (1967). The methodology of evaluation. 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>

Shute, V. J., Dennen, V. P., Kim, Y., Donmez, O., & Wang, C. (2010). *21st century assessment to promote 21st century learning: The benefits of blinking*.

Shute, V. J., & Rahimi, S. (2021). Stealth assessment of creativity in a physics video game. *Computers in Human Behavior*, *116*, 106647. <https://doi.org/10.1016/j.chb.2020.106647>

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

Smith Glasgow, M. E., Dreher, H. M., & Schreiber, J. (2019). Standardized testing in nursing education: Preparing students for NCLEX-RN and practice. *Journal of Professional Nursing*, *35*(6), 440–446. <https://doi.org/ggqh9p>

Stadler, M., Hofer, S., & Greiff, S. (2020). First among equals: Log data indicates ability differences despite equal scores. *Computers in Human Behavior*, *111*, 106442. <https://doi.org/10.1016/j.chb.2020.106442>

*Standards for Educational and Psychological Testing*. (2014). American Educational Research Association.

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.

Watters, Audrey. (2015). Speak & Spell: A History. In *Hack Education*. http://hackeducation.com/2015/01/13/speak-and-spell.

Watters, A. (2021). *Teaching machines: The history of personalized learning*. MIT 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>

Woldeab, D., & Brothen, T. (2019). 21st Century assessment: Online proctoring, test anxiety, and student performance. *International Journal of E-Learning & Distance Education*, *34*(1).

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149–162. <https://doi.org/f8txgm>