Approaches to Assessment in Higher Education

Colin Madland

2022-03-05

Table of Contents

# Introduction

Assessing learning is both a critical component of the work of teaching in higher education (HE) and also a major factor in learners’ experiences of HE ([J. Biggs 1999](#ref-biggsWhatStudentDoes1999); [Menucha Birenbaum 2007](#Xed666ae61dbae1b52e6dfe20fa7c7ad7b3c36ec); [Woldeab and Brothen 2019](#ref-woldeab21stCenturyAssessment2019)). Instructors in HE often inherit the assessment practices of their supervisors and instructors ([Lipnevich et al. 2020](#ref-lipnevichWhatGradesMean2020)), potentially leading to over-reliance on practices that may not align with modern assessment theories ([Boud 2020](#ref-boudChallengesReformingHigher2020); [Massey, DeLuca, and LaPointe-McEwan 2020](#ref-masseyAssessmentLiteracyCollege2020); [Shepard 2000](#ref-shepardRoleAssessmentLearning2000)). These traditional practices tend to emphasize summative, selected-response tests ([Flores et al. 2015](#X7f237bfd32991f9eb9aea8c786ccf0bd0c69b22); [Harlen and Deakin Crick 2002](#ref-harlenSystematicReviewImpact2002); [Lipnevich et al. 2020](#ref-lipnevichWhatGradesMean2020)) which may be characterized by low reliability, leading to inferences which may not be valid reflections of learner achievement ([Knight 2002](#ref-knightSummativeAssessmentHigher2002); [Smith Glasgow, Dreher, and Schreiber 2019](#X88b1c0c95c5ecb040b63ac0eab37de36c7ef351)). Also, the influence of an increasingly technological society is pushing higher education institutions (HEI) to integrate digital tools into the teaching and learning process ([Pellegrino and Quellmalz 2010](#Xd3220b3c4c74762840488f1ab7714e7e895bc2e)). This trend saw a massive increase in activity as a result of the COVID-19 pandemic that forced most HEI in Canada to pivot to emergency remote teaching in the spring of 2020 through at least the summer of 2021 and perhaps beyond. While this increased use of technology may be understood by some as progress, research suggests that digital technologies typically are used in ways which mimic and re-inscribe traditional assessment practices and simply increase the efficiency with which they can be administered ([Broadfoot 2016](#X51a879b1b801bbc5ad6e7fc588f4456f2973b2d)). Additionally, the needs of 21st century employers, who are looking for employees with demonstrated ability in collaboration, creative problem-solving, analytical thinking, and the ability to learn, competencies not easily measured in traditional testing formats, as opposed to those who have simply demonstrated high levels of domain-specific knowledge, have changed ([Forum 2020](#Xaf0f34faf6db1662e8da0b9850d3b9cfe5dbeed)). Finally, assessment practices have significant effects on learners including anxiety about high-stakes testing and their prospects for future studies and employment ([Harlen and Deakin Crick 2002](#ref-harlenSystematicReviewImpact2002)), and also the approaches they take to learning activities ([J. Biggs and Tang 2011](#ref-biggsTeachingQualityLearning2011); [DeLuca, Coombs, et al. 2019](#ref-delucaDifferentialSituatedView2019)). The confluence of these influences suggests that traditional approaches to assessment in HEI are becoming inadequate to meet the needs of modern HEI and ought to be reconsidered.

## Topic of the Research

A starting point for this process of examining assessment in higher education ought to be intentionally considered as 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)). Assessment practices are highly resistant to change, in part, as Broadfoot claims, because they are 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 and Wiliam 1998](#ref-blackAssessmentClassroomLearning1998); [DeLuca, Coombs, et al. 2019](#ref-delucaDifferentialSituatedView2019); [Stiggins 1991](#ref-stigginsAssessmentLiteracy1991); [Willis, Adie, and Klenowski 2013](#X731be6e2e0895c79e0f034fc829fd623ca4b0d3)). Accordingly, it is incumbent upon researchers who wish to influence assessment practice to begin with a nuanced investigation of how HE instructors approach their assessment practice and what are the impacts of those practices on learners.

## Problem to be Researched

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 and Skinner 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 (who was actually 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. 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)). Consequently, 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 ([Crooks 1988](#ref-crooksImpactClassroomEvaluation1988)).

## Purpose of the Research

Following previous research by DeLuca and colleagues ([DeLuca et al. 2016](#X773f46dba91e2266ea33fc285a547bc86776461); [DeLuca, Rickey, and Coombs 2021](#Xcffd07f338ae9c1e0f3a4638f84c8ddd08fd0be); [DeLuca, LaPointe-McEwan, and Luhanga 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, Rickey, and Coombs 2021](#Xcffd07f338ae9c1e0f3a4638f84c8ddd08fd0be); [Offerdahl and 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

This research will form the basis for an ongoing research agenda focused on research-supported strategies for assessment reform in Canadian HEI. Understanding both how instructors think about assessment and how learners are impacted by assessment decisions will be critical to informing assessment practices and policies as higher education emerges from the pandemic and moves forward into the 21st century.

# 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, Coombs, et al. 2019](#ref-delucaDifferentialSituatedView2019)) as it relates to the approaches to assessment ([DeLuca, LaPointe-McEwan, and Luhanga 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, 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, 61](#ref-blackAssessmentClassroomLearning1998))), reliance on summative assessments in HE has remained high Lipnevich et al. ([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, Chudowsky, and Glaser 2001, 112](#ref-pellegrinoKnowingWhatStudents2001)) or, more simply, “reasoning from evidence” ([Pellegrino, Chudowsky, and Glaser 2001, 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, 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 ([1996](#X6d502c1fbcba67bdb61d23bc33194bf5c0f1ecb), [1999](#ref-biggsWhatStudentDoes1999)).

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, DeLuca, and LaPointe-McEwan 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, Lake, and Matters 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, Lake, and Matters 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, DeLuca, and LaPointe-McEwan 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, DeLuca, and LaPointe-McEwan ([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, and Luhanga 2016a](#Xedee9bc10fe7ddda531b14b05542852ae9f5f1e)), “a basic understanding of educational assessment and related skills to apply such knowledge to various measures of student achievement” ([Xu and 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, Adie, and Klenowski 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, NCME, and NEA 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, 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, and Luhanga 2016b](#ref-delucaTeacherAssessmentLiteracy2016); [Shepard 2000](#ref-shepardRoleAssessmentLearning2000); [Xu and 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 and Andrade 2019](#ref-pastoreTeacherAssessmentLiteracy2019); [Xu and 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, DeLuca, and LaPointe-McEwan 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, Coombs, et al. 2019](#ref-delucaDifferentialSituatedView2019); [Willis, Adie, and Klenowski 2013](#X731be6e2e0895c79e0f034fc829fd623ca4b0d3)).

The themes DeLuca et al ([2021, 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

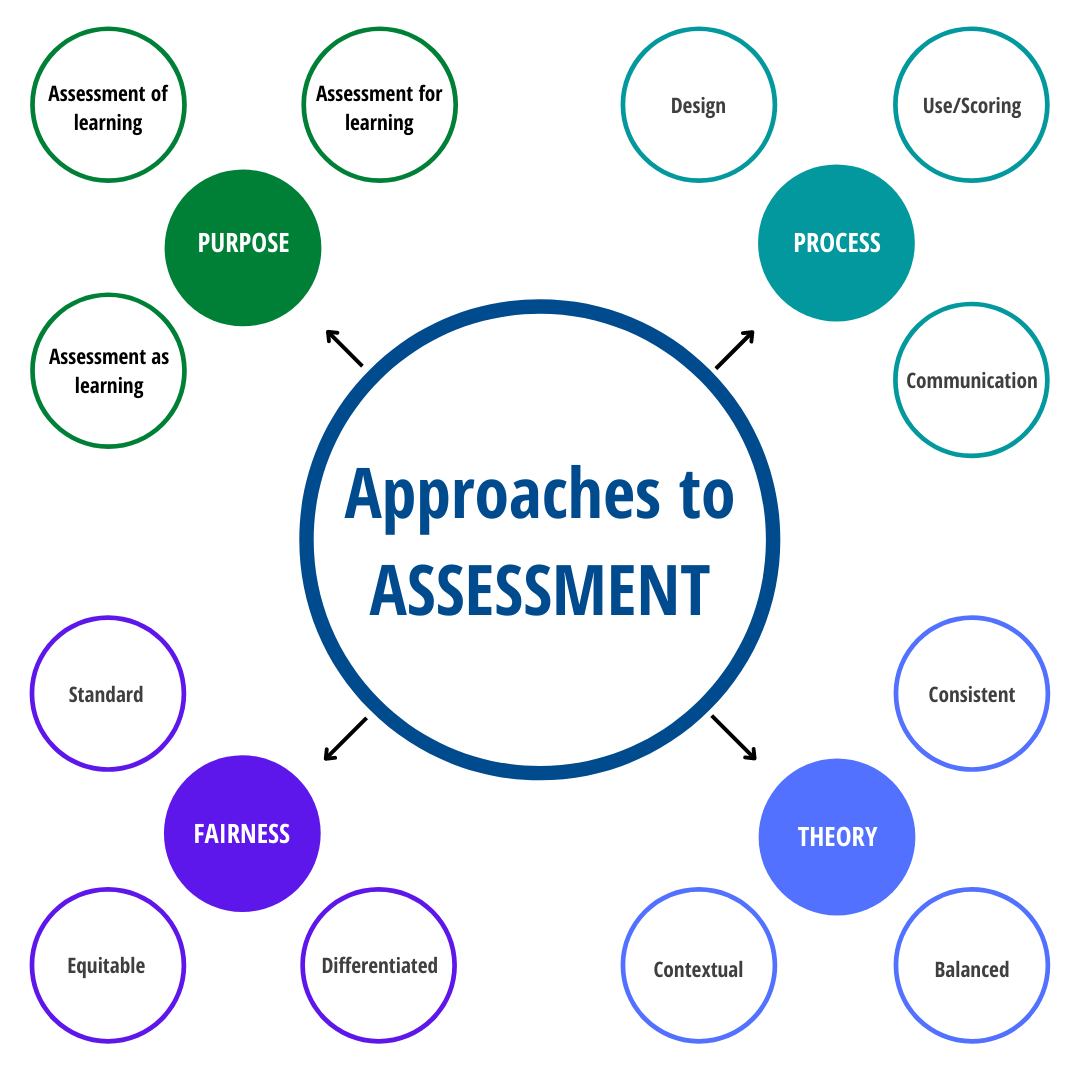
* Design
  + Focuses on the development of reliable assessments and items that measure student learning in relation to learning objectives.
* Use/scoring
  + Focuses of the adjustment and use of scoring protocols and grading schemes to respond to assessment scenarios.
* Communication
  + focuses on the 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, Rickey, and Coombs 2021, 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 and 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, Dreher, and Schreiber 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, Dreher, and Schreiber 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, 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 and Dede 2010](#X8d225d8bf3a6cb9ee5ffa78c73dfa381c6a7c30); [DeLuca, Rickey, and Coombs 2021](#Xcffd07f338ae9c1e0f3a4638f84c8ddd08fd0be); [Pellegrino and 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 and Quellmalz 2010](#Xd3220b3c4c74762840488f1ab7714e7e895bc2e); [Webb and Ifenthaler 2018](#ref-webbAssessmentTwentyFirstCentury2018)).

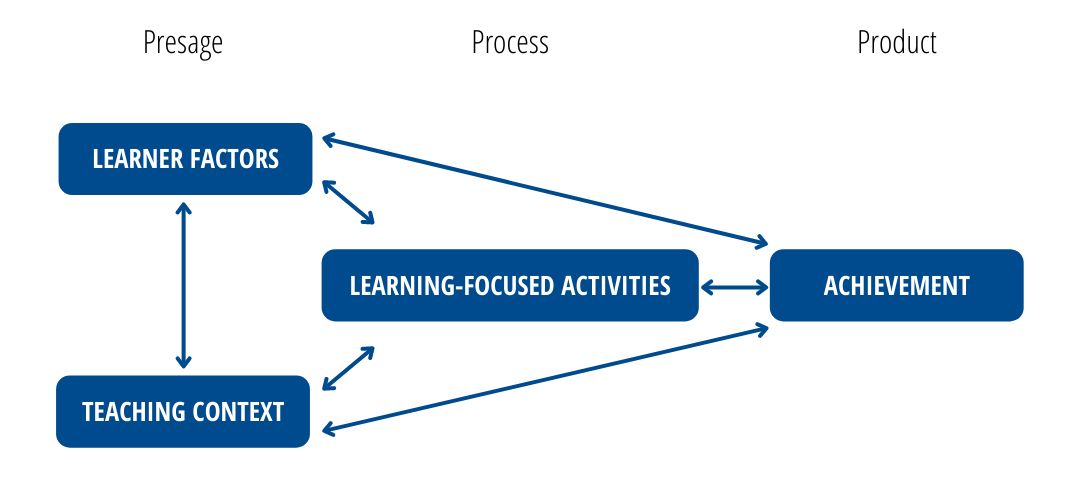
## Impact on Learners

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 and 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, and Luhanga 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))

# Validating the ACAI in Online higher Education

## Overview of the Problem, Purpose, and Questions

Research suggests that higher education faculty receive little formal preparation for assessing learning ([Lipnevich et al. 2020](#ref-lipnevichWhatGradesMean2020)) and that there may be misalignment between modern pedagogical practices and the approaches to assessment taken by higher education instructors ([Knight 2002](#ref-knightSummativeAssessmentHigher2002); [Massey, DeLuca, and LaPointe-McEwan 2020](#ref-masseyAssessmentLiteracyCollege2020); [Shepard 2000](#ref-shepardRoleAssessmentLearning2000)). The purpose of this paper will be to outline a methodological approach to investigating higher education instructors’ assessment literacies and practices as well as the impact of those practices on learners. Assessment literacy is the result of a complex interplay of skills, beliefs, and values about four domains of assessment practice: (1) the purposes of assessment; (2) the processes associated with designing, administering, scoring, and communicating the results of assessment tasks; (3) fairness in assessment practices; and (4) assessment theory, each with three priority areas. ([DeLuca, Coombs, et al. 2019](#ref-delucaDifferentialSituatedView2019)).

Specific research questions to be addressed are:

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 instructors’ assessment approaches affect learners’ experiences?

In reviewing the extant literature concerning these topics and questions, and how researchers approach them, I searched the Educational Resource Information Center (ERIC), Google Scholar, and the University of Victoria Library Summon 2.0 databases. I also searched specific journals, such as Frontiers in Education, International Review of Research in Open and Distributed Learning, the Journal of Mixed Methods Research, Educational Measurement: Issues and Practice, and Assessment and Evaluation in Higher Education. Search terms included “assessment literacy”, “assessment”, “online”, “higher education”, “learning”, and “learner impact” in various combinations. I also used [researchrabbit.ai](https://researchrabbit.ai) to assist with tracking citations of seminal and other important articles and for visualizing the relationships between found articles.

## Methodological Approaches

All research is grounded in a particular tradition or worldview, often called a *paradigm*. A research paradigm is a set of philosophical assumptions and values about the universe and the ways in which we can learn and know about it ([Johnson and Christensen 2017](#Xa4c49e3464cfac336b39022a66be02a09c2ed7e)). According to Johnson and Christenson ([2017](#Xa4c49e3464cfac336b39022a66be02a09c2ed7e)), research paradigms are generally concerned with the answers to questions related to: - how we learn about something (methodology), - how we know something is true (epistemology) - how we know reality, or if something exists (ontology), - how we know if something is valuable or ethical (axiology), and; - how we communicate and present arguments (rhetoric).

The three primary research paradigms are *quantitative*, *qualitative*, and *mixed* research (often called *mixed methods* research). Quantitative researchers traditionally followed the epistemological position of *positivism*, characterized by the idea that knowledge is singular, objective, and can be discovered through the empirical analysis of numerical data ([Held 2019](#Xb8f3a9a2f170bfdd5d9f8060668a5719d9f5be2)), although most modern quantitative researchers take a *post-positivist* view which recognizes that the values and beliefs of the researcher play a role in research and challenge the idea of pure objectivity ([Tashakkori, Johnson, and Teddlie 2020](#X8daae7541e4c079ce301a7c9a89f7da03d2db6f)). Quantitative research is often confirmatory in nature, with the researcher gathering data to either confirm or disconfirm a specific hypothesis ([Johnson and Christensen 2017](#Xa4c49e3464cfac336b39022a66be02a09c2ed7e)). Qualitative researchers, on the other hand, tend to take the epistemological view of *constructivism* which posits that meaning is constructed by the researcher as they observe individuals and groups in relation to each other and particular phenomena. Qualitative research is exploratory in nature and uses non-numerical data such as texts, transcripts, images, and categories ([Johnson and Christensen 2017](#Xa4c49e3464cfac336b39022a66be02a09c2ed7e)) in response to open-ended questions ([Tashakkori, Johnson, and Teddlie 2020](#X8daae7541e4c079ce301a7c9a89f7da03d2db6f)). During the latter half of the 20th century, when qualitative approaches were becoming more popular the research community tended towards a dualistic, ‘either/or’ view of these two paradigms ([Niglas 2010](#X7fc34501dcad97870adec0d2e3aefa7977be738); [Tashakkori, Johnson, and Teddlie 2020](#X8daae7541e4c079ce301a7c9a89f7da03d2db6f)) to the extent that the two paradigms were seen to be fundamentally incompatible. For example, Guba ([1990, 81](#ref-gubaParadigmDialog1990) as quoted in; [Johnson and Onwuegbuzie 2004, 14](#ref-johnsonMixedMethodsResearch2004)) emphasized “accommodation between paradigms is impossible … we are led to vastly diverse, disparate, and totally antithetical ends”. Early in the 21st century, however, Johnson and Onwuegbuzie countered that argument and proposed that paradigms are better visualized on a continuum, that both paradigms are useful, and that mixing methods can “draw from the strengths and minimize the weaknesses of both in single research studies and across studies” ([2004, 13–14](#ref-johnsonMixedMethodsResearch2004)). So mixed research is characterized by an epistemologically *pragmatic* view that both quantitative and qualitative approaches are valuable and that combining numerical and non-numerical data, analyses, and results can lead to deeper understanding ([Bazeley 2018](#ref-bazeleyIntegratingAnalysesMixed2018); [Tashakkori, Johnson, and Teddlie 2020](#X8daae7541e4c079ce301a7c9a89f7da03d2db6f)). More specifically, Johnson, et al. ([2007](#ref-johnsonDefinitionMixedMethods2007)) provide the following definition of mixed research:

Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration. (p. 123)

I will use the term *mixed research* in alignment with Johnson and Christensen ([2017](#Xa4c49e3464cfac336b39022a66be02a09c2ed7e)) who advocate for that term rather than *mixed methods research* because it is more than the method of research that is mixed. Indeed, mixing occurs at the level of paradigms, questions, methods, analyses, and reporting of results ([Bazeley 2018](#ref-bazeleyIntegratingAnalysesMixed2018)).

Niglas ([2010](#X7fc34501dcad97870adec0d2e3aefa7977be738)) contends that while the epistemological and philosophical stances of the researcher are important in informing the method of research, the more important considerations are the problem, purpose and questions the researcher intends to investigate. Similarly, Bazeley ([2018](#ref-bazeleyIntegratingAnalysesMixed2018)) argues that the most important factor that researchers should consider when choosing a methodology is “whether the methods chosen and the strategies used … serve the *purpose* of the research” (p. 8, emphasis in original). Following Bazeley’s argument, and considering the purpose of this research project, a mixed approach is justified. Further justification is indicated in the diversity of approaches taken by researchers investigating assessment literacy or the impact of assessment on learners using quantitative ([DeLuca et al. 2016](#X773f46dba91e2266ea33fc285a547bc86776461); [DeLuca, Rickey, and Coombs 2021](#Xcffd07f338ae9c1e0f3a4638f84c8ddd08fd0be); [Massey, DeLuca, and LaPointe-McEwan 2020](#ref-masseyAssessmentLiteracyCollege2020); [Nayagi N and Rajendran 2020](#X2e9d1b0270cc3e67622e0de93917b7ce91abfb6); [Pereira et al. 2021](#ref-pereiraHowUndergraduatesPerceive2021)), qualitative ([Boud and Dawson 2021](#ref-boudWhatFeedbackLiterate2021); [Coombs, Ge, and DeLuca 2020](#ref-coombsSeaSeaCanadian2020); [DeLuca et al. 2021](#ref-delucaPedagogySlowSignificant2021); [Earle 2020](#ref-earleBalancingDemandsValidity2020); [Fives and Barnes 2020](#ref-fivesNavigatingComplexCognitive2020); [Medland 2019](#X24013df94dd1bbf663d248dde1bf7cb672f12ec); [Watson et al. 2017](#ref-watsonSmallDataOnline2017)), and mixed ([DeLuca et al. 2018](#X07c7affd66fd5d0adbce197f0384e7e89f12100); [Nicholson 2018](#X2a2fdb822679cb457d03d450d5e80df0d02eddb); [Iannone, Czichowsky, and Ruf 2020](#ref-iannoneImpactHighStakes2020); [Tekir 2021](#ref-tekirAlignmentIntendedEnacted2021)) approaches.

### Quantitative approaches

The construct of assessment literacy has a history of being grounded in sets of standards published by researchers, governments, or regulatory agencies and of being assessed through the collection and analysis of numeric and categorical data ([DeLuca et al. 2016](#X773f46dba91e2266ea33fc285a547bc86776461); [Gotch and French 2014](#ref-gotchSystematicReviewAssessment2014)). Gotch and French ([2014](#ref-gotchSystematicReviewAssessment2014)), in their systematic review of the literature on assessment literacy inventories, found 36 instruments, all but one of which ([Jarr 2012](#X2a9d46c1f6093f71abf8d69f6e33d54cbf46084)) gathered only quantitative (numeric and categorical) data. Studies based on these instruments (for example, Alkharusi, et al. ([2012](#Xbba0a94855ee53d1a354f9bd150ed2bbefe54df)) based on Plake, et al. ([2005](#Xc5689adf9e7ca8c04511cb4ee64c8c4cc5ddf8e))) generally included quantitative analyses because the data gathered are quantitative. Early instruments ([Mertler and Campbell 2005](#Xfbd32ce8d0c6cca37cc61f1c91c0db48fed46a2); [Barbara S. Plake 1993](#ref-plakeTeacherAssessmentLiteracy1993)) were based on the 1990 *Standards for Teacher Competency in Educational Assessment of Students* ([AFT, NCME, and NEA 1990](#ref-StandardsTeacherCompetence1990)). Since then, Brookhart ([2011](#X36e21314cdda5af289172df8b2fdcbf83ba9ea4)) published a critique of the 1990 standards arguing that they were out of date and not in alignment with modern pedagogy. This was followed a few years later by Gotch and French ([2014](#ref-gotchSystematicReviewAssessment2014)) publishing the aforementioned systematic review, during which they found that the psychometric properties of the 36 instruments published between 1991 and 2012 were low. Finally, Klinger et al. ([2015](#X29739ab7fce6a0c61ad93e769577d0974c2bf34)) published the *Classroom Assessment Standards for PreK-12 Teachers* ([Klinger et al. 2015](#X29739ab7fce6a0c61ad93e769577d0974c2bf34)). These cumulative advances led to DeLuca et al. ([2016b](#ref-delucaTeacherAssessmentLiteracy2016)) to develop the *Approaches to Classroom Assessment Inventory (ACAI)* and enabled researchers to base their quantitative investigations on an instrument shown to produce valid results in relation to a modern model of assessment literacy.

There are numerous examples of research teams using quantitative approaches to investigate assessment literacy, some of which are mentioned here. DeLuca et al. ([2016](#X773f46dba91e2266ea33fc285a547bc86776461)) surveyed 404 K-12 teachers in Canada and the USA using DeLuca et al.’s ([2016a](#Xedee9bc10fe7ddda531b14b05542852ae9f5f1e)) *Approaches to Classroom Assessment Inventory (ACAI)*, an instrument based on the 2015 Standards. DeLuca et al. ([2021](#Xcffd07f338ae9c1e0f3a4638f84c8ddd08fd0be)), recently used the same inventory in an international study comparing assessment practices in the USA, Canada, and China. Nayagi and Rajendran ([2020](#X2e9d1b0270cc3e67622e0de93917b7ce91abfb6)) also used the ACAI to investigate approaches to assessment, although they modified it for their context of pre-service teachers in India. Massey et al. ([2020](#ref-masseyAssessmentLiteracyCollege2020)) surveyed higher education instructors to investigate their approaches to and confidence with assessment. Lastly, with respect to the impact of instructors’ assessment practices on learners, Pereira et al. ([2021](#ref-pereiraHowUndergraduatesPerceive2021)), Pereira et al. ([2017](#X278251e0a83fa345cc114b9dd1fa371f66b53da)), and Flores et al. ([2015](#X7f237bfd32991f9eb9aea8c786ccf0bd0c69b22)) used surveys to collect quantitative data about the perceptions of Portuguese undergraduate learners regarding the assessment practices of their instructors. While there are more studies of assessment literacy in the K-12 sector ([Medland 2019](#X24013df94dd1bbf663d248dde1bf7cb672f12ec)), there is a growing number in the higher education sector (e.g. ([Massey, DeLuca, and LaPointe-McEwan 2020](#ref-masseyAssessmentLiteracyCollege2020); [Nayagi N and Rajendran 2020](#X2e9d1b0270cc3e67622e0de93917b7ce91abfb6); [Pereira et al. 2021](#ref-pereiraHowUndergraduatesPerceive2021))).

Research questions 1 and 2 in the present study are intended to answer questions about the assessment literacies of higher education instructors in Canada, a construct which has been shown to be amenable to quantitative survey approaches in diverse settings, including K-12 teachers, pre-service teachers, and higher education instructors both in Canada and internationally.

### Qualitative approaches

Assessment literacy has been the focus of significant attention in quantitative research, but it is also becoming clear that the construct is much deeper and more complex than models based on sets of standards and competencies ([DeLuca, Willis, et al. 2019](#ref-delucaPoliciesProgramsPractices2019), [2019](#ref-delucaPoliciesProgramsPractices2019); [Willis, Adie, and Klenowski 2013](#X731be6e2e0895c79e0f034fc829fd623ca4b0d3); [Xu and Brown 2016](#ref-xuTeacherAssessmentLiteracy2016)). Willis ([2013](#X731be6e2e0895c79e0f034fc829fd623ca4b0d3)) framed assessment literacy within multiple “horizontal and vertical structures of educational discourse” following Bernstein ([1999](#X4d78eac5dad262ef2a233c8ad9b66b0f5b3cd1c)), who described vertical discourses as structured disciplinary knowledge passed down from specialists, and horizontal discourses as more contextualized knowledge practiced and passed along at a local level. Following on this idea, DeLuca et al. ([2019](#ref-delucaDifferentialSituatedView2019)) describe assessment literacy as being a “situated and differential practice predicated on negotiated knowledges” (p. 7) and elsewhere, DeLuca et al. ([2019](#ref-delucaPoliciesProgramsPractices2019)) argue that “Assessment capability involves situated professional judgement, that is the ability to draw on learning and assessment theories and experiences to purposefully design, interpret, and use a range of assessment evidence in the service of student learning” (p. 2). Similarly, Xu and Brown’s ([2016](#ref-xuTeacherAssessmentLiteracy2016)) model, *Teacher Assessment Literacy in Practice (TALiP)*, explicitly includes socio-cultural factors, teacher judgement, and teacher identity.

The methodological implication of this shift to a more complex understanding of assessment literacy is that there seems to have been an attendant shift in researchers engaging in qualitative investigations of assessment literacy ([Bearman et al. 2017](#ref-bearmanHowUniversityTeachers2017); [DeLuca et al. 2021](#ref-delucaPedagogySlowSignificant2021); [Fives and Barnes 2020](#ref-fivesNavigatingComplexCognitive2020); [Medland 2019](#X24013df94dd1bbf663d248dde1bf7cb672f12ec); [Watson et al. 2017](#ref-watsonSmallDataOnline2017)). Qualitative methodologies exist because of the view that, according to Guba and Lincoln ([1989](#ref-gubaFourthGenerationEvaluation1989)), reality is constructed through social processes, and is not simply observed in the world. In my view, this epistemological stance within the community of qualitative researchers aligns well with the nature of assessment literacy as a phenomenon impacted by horizontal discourses. In my own review of the literature on assessment literacy, there seems to be a marked increase in the number of qualitative studies published since 2016, which may reflect a deepening understanding of the characteristics of assessment literacy.

While the trend to use qualitative approaches to investigate assessment literacy and the impact of assessment on learners is more recent compared to quantitative approaches, there are good examples of qualitative studies, some of which I briefly describe here. Watson et al. ([2017](#ref-watsonSmallDataOnline2017)) published a case study of how instructor assessment practices impacted a single learner who struggled through a graduate-level learning module. Benediktsson and Ragnarsdóttir ([2020](#X9e5ea49dd0d5d230443ad609e316cd93e10a84e)) used focus group interviews followed by semi-structured interviews with individual learners to investigate the assessment experiences of immigrant learners in Icelandic universities. Bearman et al. ([2017](#ref-bearmanHowUniversityTeachers2017)) interviewed thirty-three higher education instructors from a variety of disciplines and institutions to investigate how they designed assessments and found that there is a need for professional development activities about assessment to have a relational focus. In the UK context of having external peers provide regulatory feedback, Medland ([2019](#X24013df94dd1bbf663d248dde1bf7cb672f12ec)) used an exploratory case study approach to investigate the assessment literacy of external examiners in UK higher education. In another case study with a single subject, Fives and Barnes ([2020](#ref-fivesNavigatingComplexCognitive2020)), recognizing the complexity of classroom assessment, investigated the assessment-related routines of an experienced teacher. Their goal was to engage in a close examination of how an experienced teacher enacted assessment practices in their classroom in order to “make explicit [the] cognitive tasks involved in their work” (p. 3). More recently, DeLuca et al. (-DeLuca et al. ([2021](#ref-delucaPedagogySlowSignificant2021))) used a case study methodology to analyze data gathered from interviews, reflections, and course work to investigate the learning trajectories of 35 pre-service teachers during their program. Each of these studies was exploratory in nature, and/or designed specifically to uncover deeper meanings than could be discerned with the use of a survey or other quantitative instrument, which would require the researcher to anticipate and direct respondents toward a limited set of possible answers.

Research questions 3 and 4 of the present study are intended to explore the qualitative and subjective experiences of instructors, what factors influence their approaches to assessment and the experiences of learners as they engage in assessment activities. Both of these questions will require responses to open-ended questions and will be exploratory in nature, aligning with a qualitative paradigm and epistemology.

### Mixed approaches

Researchers have argued that using mixed approaches to research is a way to strengthen and increase confidence in the validity of results by maximizing the strengths of each paradigm while minimizing the weaknesses ([Creswell 2009](#X15e33d0567975f2cdc7cfc47dc9f1173f938850); [McKim 2017](#ref-mckimValueMixedMethods2017)). Creamer ([2018](#X19c097a02df9fa745c62b1a0ce793f7d18e0b41)) argues that mixing of qualitative and quantitative methods is the defining characteristic of mixed research. She defines mixing as “*the linking, merging, or embedding of qualitative and quantitative strands of a mixed methods study*” (p. 6, emphasis in original). In a departure from the historical paradigm wars between quantitative and qualitative approaches, Niglas ([2010](#X7fc34501dcad97870adec0d2e3aefa7977be738)) contends for the idea that mixed research studies fall on a continuum between qualitative and quantitative approaches with varying degrees of epistemological stances and methods. Given the abundance of literature exploring instructor assessment literacy and practices and the impact of those practices on learners from both quantitative and qualitative paradigms, it is not surprising that there is also a body of literature investigating these constructs using mixed research approaches ([Esfandiari et al. 2016](#Xfda91508ee188109dcf716fe680285d6d978b80); [Ogan‐Bekiroglu and Suzuk 2014](#ref-ogan-bekirogluPreServiceTeachers2014); [Solomonidou and Michaelides 2017](#X78ed7c843a0a65622e4c7f10be381d21bc08632)). Creamer ([2018](#X19c097a02df9fa745c62b1a0ce793f7d18e0b41)) suggests evaluating mixed research projects using a table to make the structure of the study more plainly visible. The following tables are overviews of selected mixed research projects investigating assessment literacy in K-12 and higher education.

**Table 1**

*Overview of* ”Students’ conceptions of assessment purposes in a low stakes secondary-school context: A mixed methodology approach”\* ([Solomonidou and Michaelides 2017](#X78ed7c843a0a65622e4c7f10be381d21bc08632))\*

| Study Component | Details | Yes/No | Comments |
| --- | --- | --- | --- |
| Rationale/Purpose for Mixing | Enhancement |  |  |
| Priority | QUAN → qual |  |  |
| Timing of Data Collection | Sequential |  |  |
| Timing of Data Analysis | Sequential |  |  |
| Mixing |  |  |  |
|  | Design | Yes | Qualitative data were used to confirm tentative inferences from quantitative analysis. |
|  | Data Collection | Yes | Semi-structured interview questions were revised in light of quantitative inferences, and participants could view their survey answers during their interviews. |
|  | Data Analysis | Yes | Blended in the report. |
|  | Inferences | Yes | Blended in the report. |
|  | Fully Integrated | Yes |  |
| Qualitative Inference | Learners value assessment information but it takes work to use the information formatively. | - | - |
| Quantitative Inference | Learners agreed more with the idea that assessment is a tool for improvement and less with the idea that assessment is negative. |  |  |
| Meta-inference | Teacher education and development should systematically include alternative assessment practices. |  | - |
| Value Added | Linking data from both surveys and interviews provided opportunity for learners to elaborate on their views and provide a richer picture of their conceptions of assessment. |  |  |

**Table 2**

*Overview of* ”A Mixed-methods, Cross-sectional Study of Assessment Literacy of Iranian University Instructors: Implications for Teachers’ Professional Development”\* ([Esfandiari et al. 2016](#Xfda91508ee188109dcf716fe680285d6d978b80))\*

| Study Component | Details | Yes/No | Comments |
| --- | --- | --- | --- |
| Rationale/Purpose for Mixing | Complementarity |  |  |
| Priority | QUAN → qual |  |  |
| Timing of Data Collection | Sequential |  |  |
| Timing of Data Analysis | Sequential |  |  |
| Mixing |  |  |  |
|  | Design | Yes | The qualitative phase was used to elaborate on the quantiative phase. |
|  | Data Collection | No |  |
|  | Data Analysis | Yes | Qualitative analysis filled gaps in quantitative inferences. |
|  | Inferences | Yes | Blended in the report. |
|  | Fully Integrated | No |  |
| Qualitative Inference | Subgroups of participants assessed learners differently, used assessment for different purposes, had different levels of training in assessment, and different views on the psychometric properties of tests. |  |  |
| Quantitative Inference | Results showed support for a three-component model of assessment literacy. |  |  |
| Meta-inference | Assessment literacy is a multi-dimensional concept affected my many factors, including teaching experience, professional development and training, and the local teaching context. |  |  |
| Value Added | Qualitative analysis was critical to helping the researchers understand the differences seen between subgroups in the quantitative analysis. |  |  |

**Table 3**

*Overview of “Pre-service teachers’ assessment literacy and its implementation into practice” (*[*Ogan‐Bekiroglu and Suzuk 2014*](#ref-ogan-bekirogluPreServiceTeachers2014)*)*

| Study Component | Details | Yes/No | Comments |
| --- | --- | --- | --- |
| Rationale/Purpose for Mixing | Triangulation and Enhancement |  |  |
| Priority | QUAN → qual+qual |  |  |
| Timing of Data Collection | Sequential |  |  |
| Timing of Data Analysis | Side-by-side comparison (concurrent) |  |  |
| Mixing |  |  |  |
|  | Design | Yes | The qualitative phase was used to validate and elaborate on the quantitative phase. |
|  | Data Collection | No |  |
|  | Data Analysis | Yes | Qualitative and quantitative data were analyzed concurrently by presenting them side-by-side for comparison. |
|  | Inferences | Yes | Blended in the report. |
|  | Fully Integrated | No |  |
| Qualitative Inference | Qualitative analysis showed alignment with quantitative analysis |  |  |
| Quantitative Inference | Pre-service teachers’ assessment literacy was determined to be “close to constructivist” (p. 352) |  |  |
| Meta-inference | Although participants were able to demonstrate knowledge of assessment literacy, they had difficulty translating that knowledge into practice. |  |  |
| Value Added | Qualitative analysis was important for determining the difficulty in translating knowledge of assessment literacy into practice of good assessment. |  |  |

**Table 4**

*Overview of “A mixed method exploration of student perceptions of assessment in nursing and biomedicine” (*[*Garvey, Hodgson, and Tighe 2021*](#ref-garveyMixedMethodExploration2021)*)*

| Study Component | Details | Yes/No | Comments |
| --- | --- | --- | --- |
| Rationale/Purpose for Mixing | Triangulation |  |  |
| Priority | QUAN+QUAL |  |  |
| Timing of Data Collection | Concurrent |  |  |
| Timing of Data Analysis | Consecutive |  |  |
| Mixing |  |  |  |
|  | Design | Yes | The qualitative phase was used to validate and elaborate on the quantitative phase. |
|  | Data Collection | Yes | Open-ended (qualitative) responses were gathered along with quantitative responses and were designed to elicit clarifying views. |
|  | Data Analysis | No |  |
|  | Inferences | Yes | Blended in the report. |
|  | Fully Integrated | No |  |
| Qualitative Inference | Qualitative analysis deeper levels of complexity compared to quantitative analysis. |  |  |
| Quantitative Inference | First-year university learners indicated positive views of assessment, regardless of their program. |  |  |
| Meta-inference | Universities should recognize that first-year learners tend to have positive views of assessment and should communicate often with them about the purposes of assessment. |  |  |
| Value Added | Qualitative analysis showed areas of misalignment with the quantitative analysis. |  |  |

Each of the preceding four studies shows different ways in which mixed research can enhance validity or provide avenues for deeper and more complex understanding of instructors’ assessment literacy and practices and the impact of those on learners. The tables illustrate the complexity of mixed studies and highlight the challenges associated with integrating quantitative and qualitative approaches at each stage of the research. However, even those studies which do not demonstrate all of the qualities of fully integrated mixed research contain valuable insights based on the integration of the two analyses.

### Epistemological Alignment with Purpose of the Research

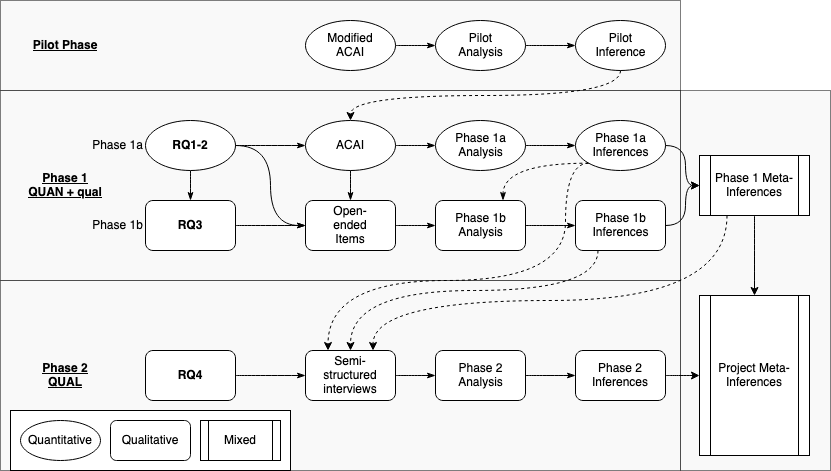
The purpose of this project is to investigate a multi-dimensional construct, assessment literacy, that is influenced by factors both internal (for example: beliefs, values, and past experiences) and external (for example: assessment culture ([DeLuca, Rickey, and Coombs 2021](#Xcffd07f338ae9c1e0f3a4638f84c8ddd08fd0be); [Massey, DeLuca, and LaPointe-McEwan 2020](#ref-masseyAssessmentLiteracyCollege2020)) or government regulations) to instructors ([DeLuca, Coombs, et al. 2019](#ref-delucaDifferentialSituatedView2019)). In addition, the project will investigate the factors that influence instructors in higher education to take particular approaches to their assessment practice and how those practices impact learners. In envisioning this multi-dimensional research project and formulating the research questions, I recognize value in both the discovery and the creation of meaning in the research process, a pragmatic epistemology that aligns with the ethos of many mixed researchers. In alignment with the examples from the literature described in the preceding sections, this project will draw from both quantitative and qualitative paradigms, will use both quantitative and qualitative data connected in multiple ways, engage in analyses in alignment with both traditions, and will present results as inferences from both quantitative and qualitative analyses in addition to meta-inferences ([Bazeley 2018, 62](#ref-bazeleyIntegratingAnalysesMixed2018)) from the integration of analyses. As such, I propose a mixed research investigation, which will be described in greater detail in the following sections.

## Visualization of a Mixed Research Project

Researchers who use mixed approaches often include visual overviews of their research project as mixed research projects can be very complex ([Creamer 2018](#X19c097a02df9fa745c62b1a0ce793f7d18e0b41)). Figure 1 below is a visual representation of the present project. It is helpful to understand some of the conventions of visualizing mixed research projects as seen in figure 1. According to Tashakkori et al. ([2020](#X8daae7541e4c079ce301a7c9a89f7da03d2db6f)), quantitative components are depicted as ovals and qualitative components are depicted as rectangles. Additionally, in figure 1, barred rectangles indicate mixed inferences. Solid arrows indicate the general direction of the workflow throughout the project and dashed arrows represent potential feedback loops where the results and inferences from one phase inform one or more components of a subsequent phase. Mixed researchers use upper- and lower-case letters to indicate the priority in any given phase. QUAN (all-caps) indicates that quantitative data and analyses are prioritized, while qual (lower-case) indicates that qualitative data and analyses are secondary. Further, a ‘+’ indicates that two or more types of data are collected and/or analyzed concurrently, while a ’ → ’ indicates that two or more types of data are collected or analyzed consecutively or sequentially. So, the notation ‘QUAN+qual’ indicates that both quantitative and qualitative data are collected and analyzed concurrently with quantitative data and analysis being prioritized. The notation ‘QUAN → QUAL’ indicates that quantitative data are being collected and/or analyzed before qualitative data, and both types of data and analysis have equal weight.

**Figure 1**

*A Visualization of a QUAN+qual → QUAL Sequential Exploratory Mixed Research Study.*



Following a pilot phase, I intend to engage in a QUAN+qual → QUAL sequential exploratory research project comprising two separate data collection and analysis phases. In phase 1a, I will use a survey of higher education instructors to collect and analyze primarily quantitative data to investigate research questions 1 and 2.

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?

This will be supplemented in phase 1b by the *concurrent collection* and *subsequent analysis* of qualitative data to answer research question 3.

1. What factors influence instructors’ approaches to assessment?

The second phase will begin following the analysis and integrationof phase 1a+b data and will collect and analyze qualitative data in semi-structured interviews with learners nominated by instructors in phase 1. Phase 2 analysis is intended to answer research question 4.

1. How do instructors’ assessment approaches affect learners’ experiences?

Since research question 4 is an integrative question ([Creamer 2018](#X19c097a02df9fa745c62b1a0ce793f7d18e0b41)), it cannot be answered fully until the inferences from phase 1 and phase 2 are integrated into one or more meta-inferences.

## Participants and Sampling Strategy

The target population for the research will be instructors who have taught at least one semester-length, accredited course in the previous 12 months at an English-speaking [^1] public higher education institution in Canada. Ideally, inferences from phase 1 of the study would be generalizable to the total population of English-speaking higher education instructors in Canada, which would require a random sample in which every member of the target population would have equal opportunity to be selected and no selection would have been influenced by a previous selection ([Hibberts, Burke Johnson, and Hudson 2012](#ref-hibbertsCommonSurveySampling2012); [Renckly 2002](#ref-rencklyAirUniversitySampling2002)). Care will be taken to ensure the sample is representative of the whole population by engaging in a stratified random sampling technique at the institutional level ([Hibberts, Burke Johnson, and Hudson 2012](#ref-hibbertsCommonSurveySampling2012)). This would involve using regions within Canada as strata (for example, British Columbia, the prairie provinces, the north, Ontario, Quebec, eastern Canada), randomly choosing 2-3 higher education institutions from within those strata, and sending the survey to all instructors at those institutions. Recruitment of participants for phase 2 will involve phase 1 participants (instructors) nominating 8-10 learners from their courses, I will randomly select 2-3 respondents from phase 1 and then randomly select 4-5 of their nominated learners to receive invitations to the semi-structured interviews. This procedure will ensure that instructors will not know which of their nominated learners participated in the research and it will create “linked data” ([2018, 126](#ref-bazeleyIntegratingAnalysesMixed2018)), allowing for greater explanatory power due to being able to match specific instructor approaches with learner impacts.

## Data Collection Strategy

### Phase 1

I propose that the phase 1 data collection instrument be DeLuca et al.’s ([2016a](#Xedee9bc10fe7ddda531b14b05542852ae9f5f1e)) ACAI (see Appendix 1 for the full ACAI v3.0 and its specifications). The ACAI was designed by a team of researchers, led by Dr. Christopher DeLuca at Queen’s University in Canada to investigate the approaches that K-12 educators take to assessment in their classrooms. It is based on an analysis of assessment standards published in five majority English-speaking countries since 1990. The instrument is based on a four-dimensional framework of assessment literacy (assessment purposes, assessment processes, assessment fairness, and measurement theory), and includes one open-ended question and several closed-ended questions about how teachers form their approaches to assessment. It also includes a section regarding educator beliefs about assessment based on the *Beliefs about Assessment* scale ([Smith et al. 2014](#ref-smithPreparingTeachersUse2014)). The ACAI has been used in a wide variety of studies within the instrument’s author’s research group ([Coombs et al. 2018](#X3310b036a234cdf230045cc2106e53d723808d6); [Coombs, DeLuca, and MacGregor 2020](#Xfe48ff53c3a3416066f58f9a41b535978fae869); [DeLuca, Coombs, et al. 2019](#ref-delucaDifferentialSituatedView2019); [DeLuca et al. 2016](#X773f46dba91e2266ea33fc285a547bc86776461); [DeLuca, Willis, et al. 2019](#ref-delucaPoliciesProgramsPractices2019); [Schneider et al. 2020](#X9107aa38217f38eb2a3cd58416130c5f29e300a)), and also by other researchers ([Nayagi N and Rajendran 2020](#X2e9d1b0270cc3e67622e0de93917b7ce91abfb6)). In several of these studies, the ACAI has been modified somewhat to fit the particular context or research design.

As the ACAI was designed for use in K-12 contexts and has been used less often to investigate higher education instructors’ approaches to assessment, I will modify the language in the instrument to be more suitable for the higher education context and to include more open-ended questions. Following these modifications, I will undertake a pilot phase to determine whether the modified ACAI is suitable for the higher education context. The pilot phase will inform the version of the ACAI used in Phase 1 of the study and will include both expert review and the recruitment of a small sample from the target population to preview the survey in its intended mode and offer specific and guided feedback ([Hibberts, Burke Johnson, and Hudson 2012](#ref-hibbertsCommonSurveySampling2012); [Renckly 2002](#ref-rencklyAirUniversitySampling2002)).

### Phase 2

The semi-structured interviews in phase 2 will include questions similar to the following:

* Please describe what you understand assessment to be.
* Did the assessment strategies that your instructor used help you to learn the course material?
* Did [instructor] use technology to administer assessments?
  + How did their use/not use of technology impact your performance on the assessments?
* Please describe your *favourite* assessment from [instructor]’s course [course name]?
  + What made that assessment your favourite?
* Please describe your *least favourite* assessment from [instructor]’s course [course name]?
  + What made that assessment your least favourite?
* What factors made it *easier* to do well on the assessments in [course]?
* What factors made it *more difficult* to do well on the assessments in [course]?
* How did the assessment strategies your instructor used make you feel?
* Do you receive any support or accommodations to allow you to complete assessments?

## Types of Data and Analysis

As shown in the visualization of the research project (Figure 1), both quantitative and qualitative data will be gathered for analysis. The following sections will describe the data I intend to gather and how it may address the research questions.

### Quantitative data

During phase 1a, quantitative data will be gathered using the ACAI and will primarily be in the form of responses to 6-point Likert scale items. Similar to Flores et al. ([2020](#X21a54516e2b8ef032e69ea54eb24962d0415dba)), the first step will be to engage in a confirmatory factor analysis to determine the fit of the data with the *Approaches to Classroom Assessment* model. This step will address research question 1. Then, as the data will be *ordinal* data (i.e. a score of five is greater than a score of three), but not *interval* data (i.e. it is not possible to determine *how much greater* a score of five is compared to a score of three) ([Bulut 2021](#ref-bulutLikertScalesFriend2021); [Jamieson 2004](#ref-jamiesonLikertScalesHow2004)), it will not be appropriate to use the data to calculate common statistics such as mean or standard deviation (i.e. there is no way to calculate the mean of “highly likely” and “highly unlikely”). Consequently, more advanced statistical techniques which rely on the mean and standard deviation (i.e. *t* test, analysis of variance), otherwise known as *parametric* tests, are not appropriate for ordinal data ([Field 2018](#ref-fieldDiscoveringStatisticsUsing2018)). Fortunately, there are non-parametric tests that enable analysis of data that do not meet the requirements for parametric tests. One example, which will be useful in comparing two different demographic groups within my data (e.g. instructors with more than two years experience teaching with technology compared to instructors with less than two years of experience teaching with technology), is the *Mann-Whitney-Wilcoxon* test ([Gao 2010](#ref-gaoNonparametricStatistics2010)) (MWW). Instead of calculating based on the mean scores of different samples, the MWW test requires the researcher to first rank the scores and then calculate the differences between the samples based on the ranked scores. If I am to compare 3 or more groups, the *Kruskal-Wallis* test ([Gao 2010](#ref-gaoNonparametricStatistics2010)), based on the same ranking procedure, would be more appropriate. These two tests will be used to address research question 2.

### Qualitative data

During phase 1b, I will gather and analyze qualitative data through the use of open-ended questions embedded in the ACAI. Part B of the ACAI currently contains open-ended questions asking respondents to describe how they would respond to the scenario if their response isn’t listed as an option in the instrument ([DeLuca ND](#ref-delucaACAIInstrumentSpecificationsND)). Below are some open-ended questions which might be added to Part A (demographics) of the ACAI:

* In a few sentences or point form notes, please describe the most important factors you consider when planning how you will assess learners in an upcoming course.
* In a few sentences or point form notes, please describe what you understand assessment to be.
* In what ways do you use digital technology in your assessment practice?

The current and proposed open-ended questions in the ACAI are intended to provide qualitative data which will be analyzed using a theory-driven thematic coding process ([DeLuca et al. 2018](#X07c7affd66fd5d0adbce197f0384e7e89f12100); [Namey et al. 2008](#ref-nameyDataReductionTechniques2008)) to explore in greater depth the factors that influence instructors to approach assessment as they do in their courses (research question 3).

Namey et al. ([2008](#ref-nameyDataReductionTechniques2008)) describe two general approaches to the analysis of qualitative data: content analysis and thematic analysis. In content analysis, the researcher “evaluates the frequency and saliency of particular words or phrases in a body of original text data in order to identify keywords or repeated ideas” (p. 138), but with the drawback that there is limited interpretation or consideration of context. In thematic analysis, researchers consider not only specific words and phrases, but also “implicit and explicit ideas” (p. 138). They further describe “data-driven” (themes emerge from the data) and “theory-driven” (the researcher identifies themes *a priori*, without consulting the data)(p. 138) approaches (see also: [Tashakkori, Johnson, and Teddlie 2020](#X8daae7541e4c079ce301a7c9a89f7da03d2db6f)). In phase 1b of the proposed project, I will take a theory-driven approach to a thematic analysis where data will be coded in alignment with the four dimensions of the *Approaches to Classroom Assessment* framework ([DeLuca et al. 2016](#X773f46dba91e2266ea33fc285a547bc86776461)) (the dotted line connecting “Phase 1a Inferences” to “Phase 1b Analysis” in figure 1). In addition, I will be particularly mindful that emergent themes may become clear in the data and I will be especially interested in those not represented in, or which are in contradiction to, the quantitative analysis ([Tashakkori, Johnson, and Teddlie 2020](#X8daae7541e4c079ce301a7c9a89f7da03d2db6f)).

After completing both phase 1a and 1b, I will be able to synthesize inferences from the two phase 1 analyses, and also one or more meta-inferences from the integrated analysis of the two sub-phases. These inferences will inform the phase 2 semi-structured interview questions, providing an integration point (the dotted lines from “Phase 1a Analysis”, “Phase 1b Analysis”, and “Phase 1 Meta-inferences” to “Semi-structured Interviews”).

Raw qualitative data from phase 2 of the project will be in the form of audio recordings of semi-structured interviews, which will be transcribed verbatim and stored as plain text files. As this data is less connected to phase 1 data because it is a different sample population, I will use a data-driven thematic approach to uncover emergent themes in the data while not precluding the possibility that theory-driven themes may emerge.

Following the phase 1 and phase 2 analyses, I will synthesize the inferences from phase 1a and 1b, the meta-inferences from phase 1, and the inferences from phase 2 to draw project-level meta-inferences from the whole of the data, thereby addressing research question 4.

## Research Procedures

Following approval from the University of Victoria Human Research Ethics Board (HREB) and a brief pilot phase, I will use a stratified random sampling technique to randomly select 2-3 higher education institutions from each of six regions in Canada (British Columbia, the prairie provinces, Ontario, Quebec, eastern Canada, and northern Canada), for a total of 12-18 institutions. I will contact the faculty associations at each of these institutions to request that they forward my letter of invitation and embedded link to a web-based survey. Each participant will be asked to self-identify and to nominate 8-10 learners in one of their courses and include the nominees’ contact information. To begin phase 2, I will randomly select 2-3 instructors from phase 1, then randomly select 4-5 of their nominated learners to invite to semi-structured interviews. The interviews will be conducted via web-conferencing software that allows recording and the automatic generation of transcripts. Table 5 shows a projected timeline for the research.

**Table 5**

***Projected Timeline***

| Projected Date | Project Stage |
| --- | --- |
| December 2021 | UVic HREB Approval |
| January 2022 | Pilot Phase |
| February 2022 | Phase 1 Data Collection |
| July 2022 | Phase 1a and 1b Data Analysis |
| September 2022 | Phase 2 Data Collection |
| November 2022 | Phase 2 Data Analysis |
| January 2023 | Integration and Write-Up |

## Chapter Summary

In this chapter, I have discussed the three major approaches to research, *quantitative*, *qualitative*, and *mixed*, along with their underlying epistemologies. Through a review of the literature on assessment literacy and its impact on learners in higher education, I show that researchers use all three approaches and that the most important consideration in determining a method to pursue an investigation in this area is the purpose of the investigation. As the purpose of this investigation is to gain understanding of the multi-dimensional construct of assessment literacy *and* its effects on learners, a highly subjective phenomenon, a mixed research approach is justified. I discuss a possible method to gather and analyze quantitative and qualitative data from higher education instructors and their learners, along with strategies for drawing inferences and meta-inferences based on the integration of the separate analyses.

[^1] If funding for translation services is secured, I would extend the sample population to include French-speaking instructors and learners.

# Assessment in Digital Learning

* Emergent themes based on first two papers, which ***may*** include:
  + technology for adaptive testing, automated item generation, large-scale assessments (NCLEX, PISA, FSA, etc)
  + technology for surveillance or academic integrity (Turnitin, Proctorio, etc)
  + technology to enable new forms of learning- and learner-centred assessments (portfolios, collaboration, version control, learner contributors)
  + ethical applications of technology in assessment

### Introduction

## 0.1 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 and 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, Dreher, and Schreiber 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, Dreher, and Schreiber 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, 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 and Dede 2010](#X8d225d8bf3a6cb9ee5ffa78c73dfa381c6a7c30); [DeLuca, Rickey, and Coombs 2021](#Xcffd07f338ae9c1e0f3a4638f84c8ddd08fd0be); [Pellegrino and 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 and Quellmalz 2010](#Xd3220b3c4c74762840488f1ab7714e7e895bc2e); [Webb and Ifenthaler 2018](#ref-webbAssessmentTwentyFirstCentury2018)).

### Literature

### Theoretical / Conceptual Framework

### Research Method

### Analysis

### Findings and Limitations

### Discussion

### Impact on Theory and/or Practice

### Recommendations for Future Research

### Conclusion

### Data Availability

### References

## Conclusions and Recommendations

## References

## Appendices

# Discussion

We have finished a nice book.

# Conclusion

This is from OSF

AFT, NCME, and NEA. 1990. “Standards for Teacher Competence in Educational Assessment of Students.” <https://eric.ed.gov/?id=ED323186>.

Alkharusi, Hussain, Said Aldhafri, Hilal Alnabhani, and Muna Alkalbani. 2012. “Educational Assessment Attitudes, Competence, Knowledge, and Practices: An Exploratory Study of Muscat Teachers in the Sultanate of Oman.” *Journal of Education and Learning* 1 (2): p217. <https://doi.org/r7f>.

Bazeley, Patricia. 2018. *Integrating Analyses in Mixed Methods Research*. 1 Oliver’s Yard, 55 City Road London EC1Y 1SP: SAGE Publications Ltd. <https://doi.org/10.4135/9781526417190>.

Bearman, Margaret, Phillip Dawson, Sue Bennett, Matt Hall, Elizabeth Molloy, David Boud, and Gordon Joughin. 2017. “How University Teachers Design Assessments: A Cross-Disciplinary Study.” *Higher Education* 74 (1): 49–64. <https://doi.org/gbhr94>.

Bearman, Margaret, Phillip Dawson, David Boud, Sue Bennett, Matt Hall, and Elizabeth Molloy. 2016. “Support for Assessment Practice: Developing the Assessment Design Decisions Framework.” *Teaching in Higher Education* 21 (5): 545–56. <https://doi.org/gjsn3g>.

Benediktsson, Artëm Ingmar, and Hanna Ragnarsdóttir. 2020. “Immigrant Students’ Experiences of Assessment Methods Used in Icelandic Universities.” *Multicultural Education Review* 12 (2): 98–116. <https://doi.org/gmcv29>.

Benjamin, Ludy T. 1988. “A History of Teaching Machines.” *American Psychologist* 43 (9): 703–12. <https://doi.org/djgzjr>.

Bernstein, Basil. 1999. “Vertical and Horizontal Discourse: An Essay.” *British Journal of Sociology of Education* 20 (2): 157–73. <https://doi.org/ftmsvc>.

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

Biggs, John. 1996. “Enhancing Teaching Through Constructive Alignment.” *Higher Education* 32 (3): 347–64. <https://doi.org/chx3gp>.

———. 1999. “What the Student Does: Teaching for Enhanced Learning.” *Higher Education Research & Development* 18 (1): 57–75. <https://doi.org/drgphk>.

Biggs, John, and Catherine Tang. 2011. *Teaching for Quality Learning at University: What the Student Does*. 4th ed. New York: Society for Research into Higher Education & Open University Press.

Birenbaum, M. 1996. “Assessment 2000: Towards a Pluralistic Approach to Assessment.” In *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge*, edited by M Birenbaum and F. J. R. C. Dochy, 42:3–29. Kluwer Academic/Plenum Publishers.

———. 2003. “New Insights into Learning and Teaching and Their Implications for Assessment.” In *Opti- Mising New Modes of Assessment: In Search of Qualities and Standards*, edited by M Segers, F Dochy, and E Cascallar, 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): 749–68. <https://doi.org/cqgvs8>.

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

Bloom, Benjamin. 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. <https://eric.ed.gov/?id=ED053419>.

Boud, David. 2020. “Challenges in Reforming Higher Education Assessment: A Perspective from Afar.” *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa* 26 (1). <https://doi.org/gjvpw5>.

Boud, David, and Phillip Dawson. 2021. “What Feedback Literate Teachers Do: An Empirically-Derived Competency Framework.” *Assessment & Evaluation in Higher Education*, April, 1–14. <https://doi.org/gjrsqx>.

Braguinski, Nikita. 2018. “An (An)Archive of Communication: Interactive Toys as Interlocutor.” *Communication +1* 7 (1): 19. <https://pure.mpg.de/rest/items/item_3326892/component/file_3327086/content>.

Broadfoot, Patricia. 2016. “Assessment for Twenty-First-Century Learning: The Challenges Ahead.” In *Learning, Design, and Technology*, edited by Michael J Spector, Barbara B Lockee, and Marcus D. Childress, 1–23. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-17727-4_64-1>.

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

Brown, Gavin. 2017. “Teachers Conceptions of Assessment - Secondary Schools Long and Abridged.” <https://doi.org/gj4tz6>.

Brown, Gavin T. L., Robert Lake, and Gabrielle Matters. 2011. “Queensland Teachers’ Conceptions of Assessment: The Impact of Policy Priorities on Teacher Attitudes.” *Teaching and Teacher Education* 27 (1): 210–20. <https://doi.org/c3k8f5>.

Bulut, Okan. 2021. “Likert Scales: Friend or Foe?” Medium. August 4, 2021. <https://towardsdatascience.com/likert-scales-friend-or-foe-76f865786fb7>.

Caputi, Linda 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, Jody, and Chris Dede. 2010. “Assessment, Technology, and Change.” *Journal of Research on Technology in Education* 42 (3): 309–28. <https://doi.org/ghdnhh>.

Coombs, Andrew, Christopher DeLuca, Danielle LaPointe-McEwan, and Agnieszka Chalas. 2018. “Changing Approaches to Classroom Assessment: An Empirical Study Across Teacher Career Stages.” *Teaching and Teacher Education* 71 (April): 134–44. <https://doi.org/gdfmcv>.

Coombs, Andrew, Christopher DeLuca, and Stephen MacGregor. 2020. “A Person-Centered Analysis of Teacher Candidates’ Approaches to Assessment.” *Teaching and Teacher Education* 87 (January): 102952. <https://doi.org/gh5k6v>.

Coombs, Andrew, Jenny Ge, and Christopher DeLuca. 2020. “From Sea to Sea: The Canadian Landscape of Assessment Education.” *Educational Research*, October, 1–17. <https://doi.org/gh5k4z>.

Creamer, Elizabeth G. 2018. *An Introduction to Fully Integrated Mixed Methods Research*. 2455 Teller Road, Thousand Oaks California 91320: SAGE Publications, Inc. <https://doi.org/10.4135/9781071802823>.

Creswell, John W. 2009. *Research Design: Qualitative, Quantitative and Mixed Method Approaches*. 3rd ed. Thousand Oaks, CA: Sage Publications.

Crooks, Terence J. 1988. “The Impact of Classroom Evaluation Practices on Students.” *Review of Educational Research* 58 (4): 438–81. <https://doi.org/dvd8nf>.

DeLuca, Christopher. ND. “ACAI Instrument, Specifications, and Related Material.” Classroom Assessment Research Team (CART).

———. 2012. “Preparing Teachers for the Age of Accountability: Toward a Framework for Assessment Education.” *Action in Teacher Education* 34: 576–91. <https://doi.org/10.1080/01626620.2012.730347>.

DeLuca, Christopher, Allison E. A. Chapman-Chin, Danielle LaPointe-McEwan, and Don A. Klinger. 2018. “Student Perspectives on Assessment for Learning.” *The Curriculum Journal* 29 (1): 77–94. <https://doi.org/gf4xvc>.

DeLuca, Christopher, Teresa Chavez, and Chunhua Cao. 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): 107–26. <https://doi.org/gj5v98>.

DeLuca, Christopher, A. Coombs, S. Macgregor, and A. Rasooli. 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, Christopher, Danielle LaPointe-McEwan, and Ulemu Luhanga. 2016a. “Approaches to Classroom Assessment Inventory: A New Instrument to Support Teacher Assessment Literacy.” *Educational Assessment* 21: 248–66. <https://doi.org/gfgtsg>.

———. 2016b. “Teacher Assessment Literacy: A Review of International Standards and Measures.” *Educational Assessment, Evaluation and Accountability* 28: 251–72. <https://doi.org/f828mh>.

DeLuca, Christopher, Nathan Rickey, and Andrew Coombs. 2021. “Exploring Assessment Across Cultures: Teachers’ Approaches to Assessment in the U.S., China, and Canada.” Edited by Sammy King Fai Hui. *Cogent Education* 8 (1): 1921903. <https://doi.org/gjxvc7>.

DeLuca, Christopher, Michelle Searle, Katrina Carbone, Jenny Ge, and Danielle LaPointe-McEwan. 2021. “Toward a Pedagogy for Slow and Significant Learning about Assessment in Teacher Education.” *Teaching and Teacher Education* 101 (May): 103316. <https://doi.org/gh7mjt>.

DeLuca, Christopher, Adelina Valiquette, A. Coombs, Danielle LaPointe-McEwan, and Ulemu Luhanga. 2016. “Teachers’ Approaches to Classroom Assessment: A Large-Scale Survey.” *Assessment in Education: Principles, Policy & Practice* 25: 355–75. <https://doi.org/gh5k6p>.

DeLuca, Christopher, Jill Willis, Bronwen Cowie, Christine Harrison, Andrew Coombs, Andrew Gibson, and Suzanne Trask. 2019. “Policies, Programs, and Practices: Exploring the Complex Dynamics of Assessment Education in Teacher Education Across Four Countries.” *Frontiers in Education* 4 (November): 132. <https://doi.org/gh5k2r>.

Earl, Lorna M. 2013. *Assessment as Learning: Using Classroom Assessment to Maximize Student Learning*. Second edition. Thousand Oaks, Calif: Corwin Press.

Earle, Sarah. 2020. “Balancing the Demands of Validity and Reliability in Practice: Case Study of a Changing System of Primary Science Summative Assessment.” *London Review of Education*. <https://doi.org/gmgbps>.

Esfandiari, Rajab, Imam Khomeini International University, Qazvin, Razieh Nouri, and Imam Khomeini International University, Qazvin. 2016. “A Mixed-Methods, Cross-Sectional Study of Assessment Literacy of Iranian University Instructors: Implications for Teachers’ Professional Development.” *Iranian Journal of Applied Linguistics* 19 (2): 115–54. <https://doi.org/gmdnd9>.

Field, Andy P. 2018. *Discovering Statistics Using IBM SPSS Statistics*. 5th edition, North American edition. Thousand Oaks, California: Sage Publications Inc.

Fives, Helenrose, and Nicole Barnes. 2020. “Navigating the Complex Cognitive Task of Classroom Assessment.” *Teaching and Teacher Education* 92 (June): 103063. <https://doi.org/gmbws5>.

Fletcher, Richard B., Luanna H. Meyer, Helen Anderson, Patricia Johnston, and Malcolm Rees. 2012. “Faculty and Students Conceptions of Assessment in Higher Education.” *Higher Education* 64 (1): 119–33. <https://doi.org/ctccpq>.

Flores, Maria Assunção, Gavin Brown, Diana Pereira, Clara Coutinho, Patrícia Santos, and Cláudia Pinheiro. 2020. “Portuguese University Students’ Conceptions of Assessment: Taking Responsibility for Achievement.” *Higher Education* 79 (3): 377–94. <https://doi.org/gk4j65>.

Flores, Maria Assunção, Ana Margarida Veiga Simão, Alexandra Barros, and Diana Pereira. 2015. “Perceptions of Effectiveness, Fairness and Feedback of Assessment Methods: A Study in Higher Education.” *Studies in Higher Education* 40 (9): 1523–34. <https://doi.org/gfz39k>.

Forum, World Economic. 2020. “The Future of Jobs Report 2020.” World Economic Forum. <https://www.weforum.org/reports/the-future-of-jobs-report-2020/>.

Frantz, Gene. 2014. *The Speak N Spell*. OpenStax CNX. <http://cnx.org/contents/b3014cd9-6e3a-474f-bbaa-c948cbc8e205@5.12.>

Gao, Xin. 2010. “Nonparametric Statistics.” In *Encyclopedia of Research Design*, edited by Neil Salkind. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc. <https://doi.org/10.4135/9781412961288.n272>.

Garvey, Loretta, Yvonne Hodgson, and Josephine Tighe. 2021. “A Mixed Method Exploration of Student Perceptions of Assessment in Nursing and Biomedicine.” *Journal of Further and Higher Education*, March, 1–14. <https://doi.org/gmb9k6>.

Gotch, Chad, and Brian F. French. 2014. “A Systematic Review of Assessment Literacy Measures.” *Educational Measurement: Issues and Practice* 33: 14–18. <https://doi.org/gcpg4p>.

Guba, E. G., and Y. S. Lincoln. 1989. *Fourth Generation Evaluation*. SAGE Publications. <https://books.google.ca/books?id=k_zxEUst46UC>.

Guba, Egon G., ed. 1990. *The Paradigm Dialog*. Newbury Park, Calif: Sage Publications.

Harlen, Wynne, and Ruth Deakin Crick. 2002. “A Systematic Review of the Impact of Summative Assessment and Tests on Students’ Motivation for Learning.” *Research Evidence in Education Library*, no. 1: 151. <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=108>.

Held, Mirjam B. E. 2019. “Decolonizing Research Paradigms in the Context of Settler Colonialism: An Unsettling, Mutual, and Collaborative Effort.” *International Journal of Qualitative Methods* 18 (January). <https://doi.org/10.1177/1609406918821574>.

Herppich, Stephanie, Anna-Katharina Praetorius, Natalie Förster, Inga Glogger-Frey, Karina Karst, Detlev Leutner, Lars Behrmann, et al. 2018. “Teachers’ Assessment Competence: Integrating Knowledge-, Process-, and Product-Oriented Approaches into a Competence-Oriented Conceptual Model.” *Teaching and Teacher Education* 76 (November): 181–93. <https://doi.org/gfjsvn>.

Hibberts, Mary, R. Burke Johnson, and Kenneth Hudson. 2012. “Common Survey Sampling Techniques.” In *Handbook of Survey Methodology for the Social Sciences*, edited by Lior Gideon, 53–74. New York, NY: Springer New York. <https://doi.org/10.1007/978-1-4614-3876-2_5>.

Iannone, Paola, Christoph Czichowsky, and Johannes Ruf. 2020. “The Impact of High Stakes Oral Performance Assessment on Students’ Approaches to Learning: A Case Study.” *Educational Studies in Mathematics* 103 (3): 313–37. <https://doi.org/gmcvbf>.

Irvine, Valerie. 2020. “The Landscape of Merging Modalities.” *EDUCAUSE Review* 2020 (4): 40–58. <https://er.educause.edu/articles/2020/10/the-landscape-of-merging-modalities>.

Jamieson, Susan. 2004. “Likert Scales: How to (Ab)use Them.” *Medical Education* 38 (12): 1217–18. <https://doi.org/b5gxwx>.

Jarr, Karoline Ann. 2012. “Education Practitioners’ Interpretation and Use of Assessment Results.” Doctor of Philosophy, University of Iowa. <https://doi.org/10.17077/etd.35vh2oc1>.

Johnson, R. Burke, and Larry Christensen. 2017. *Educational Research: Quantitative, Qualitative and Mixed Approaches*. 6th ed. Thousand Oaks, CA: Sage Publications.

Johnson, R. Burke, and Anthony J. Onwuegbuzie. 2004. “Mixed Methods Research: A Research Paradigm Whose Time Has Come.” *Educational Researcher* 33 (7): 14–26. <https://doi.org/cg679w>.

Johnson, R. Burke, Anthony J. Onwuegbuzie, and Lisa A. Turner. 2007. “Toward a Definition of Mixed Methods Research.” *Journal of Mixed Methods Research* 1 (April): 112–33. <https://doi.org/10.1177/1558689806298224>.

Jones, Emma, Michael Priestley, Liz Brewster, Susan J. Wilbraham, Gareth Hughes, and Leigh Spanner. 2021. “Student Wellbeing and Assessment in Higher Education: The Balancing Act.” *Assessment & Evaluation in Higher Education* 46 (3): 438–50. <https://doi.org/gk36pd>.

Klinger, Don, Patricia McDivitt, Barbara Howard, Todd Rogers, Marco Munoz, and Caroline Wylie. 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>.

Knight, Peter T. 2002. “Summative Assessment in Higher Education: Practices in Disarray.” *Studies in Higher Education* 27 (3): 275–86. <https://doi.org/b25nb2>.

Lin, Qiao, Yue Yin, Xiaodan Tang, Roxana Hadad, and Xiaoming Zhai. 2020. “Assessing Learning in Technology-Rich Maker Activities: A Systematic Review of Empirical Research.” *Computers & Education* 157 (November): 103944. <https://doi.org/ghgzp9>.

Lipnevich, Anastasiya A., Thomas R. Guskey, Dana M. Murano, and Jeffrey K. Smith. 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, Kyle D., Christopher DeLuca, and Danielle LaPointe-McEwan. 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>.

McKim, Courtney A. 2017. “The Value of Mixed Methods Research: A Mixed Methods Study.” *Journal of Mixed Methods Research* 11 (2): 202–22. <https://doi.org/f952f7>.

Medland, Emma. 2015. “Examining the Assessment Literacy of External Examiners.” *London Review of Education*, December. <https://doi.org/gk5sph>.

———. 2019. “‘I’m an Assessment Illiterate’: Towards a Shared Discourse of Assessment Literacy for External Examiners.” *Assessment & Evaluation in Higher Education* 44 (4): 565–80. <https://doi.org/gj47j3>.

Mertler, Craig A, and Cynthia Campbell. 2005. “Measuring Teachers’ Knowledge & Application of Classroom Assessment Concepts: Development of the "Assessment Literacy Inventory".” Presented at the Annual Meeting of the American Educational Research Association, Montreal, QC, CA, April 11–15. <https://eric.ed.gov/?id=ED490355>.

Mislevy, Robert J. 1994. “Test Theory Reconcieved.” *ETS Research Report Series* 1994 (1): i–38. <https://doi.org/gjm236>.

Namey, Emily, Greg Guest, Lucy Thairu, and Laura Johnson. 2008. “Data Reduction Techniques for Large Qualitative Data Sets.” In *Handbook for Team-Based Qualitative Research*, 137–62.

Natriello, Gary. 1987. “The Impact of Evaluation Processes on Students.” *Educational Psychologist* 22 (2): 155–75. <https://doi.org/cgqtqx>.

Nayagi N, Kothai, and M. Rajendran. 2020. “Pre-Service Teachers’ Approaches to Assessment.” *Humanities & Social Sciences Reviews* 8 (1): 666–73. <https://doi.org/10.18510/hssr.2020.8180>.

Nicholson, Dawn Theresa. 2018. “Enhancing Student Engagement Through Online Portfolio Assessment.” *Practitioner Research in Higher Education* 11 (1): 15–31.

Niglas, Katrin. 2010. “The Multidimensional Model of Research Methodology: An Integrated Set of Continua.” In *SAGE Handbook of Mixed Methods in Social & Behavioral Research*, edited by Abbas Tashakkori and Charles Teddlie, 215–36. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc. <http://methods.sagepub.com/book/sage-handbook-of-mixed-methods-social-behavioral-research-2e/n9.xml>.

Offerdahl, Erika G., and Debra Tomanek. 2011. “Changes in Instructors’ Assessment Thinking Related to Experimentation with New Strategies.” *Assessment and Evaluation in Higher Education* 36 (7): 781–95. <https://doi.org/d89z7p>.

Ogan‐Bekiroglu, Feral, and Erol Suzuk. 2014. “Pre‐service Teachers’ Assessment Literacy and Its Implementation into Practice.” *The Curriculum Journal* 25 (3): 344–71. <https://doi.org/gmdnfr>.

Oldfield, Alison, Patricia Broadfoot, Rosamund Sutherland, and Sue Timmis. 2012. “Assessment in a Digital Age: A Research Review.” Graduate School of Education, University of Bristol. <https://www.bristol.ac.uk/media-library/sites/education/documents/researchreview.pdf>.

Pastore, Serafina, and Heidi L. Andrade. 2019. “Teacher Assessment Literacy: A Three-Dimensional Model.” *Teaching and Teacher Education* 84 (August): 128–38. <https://doi.org/gh5k7b>.

Pellegrino, James W., Naomi Chudowsky, and Robert Glaser. 2001. *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, D.C.: National Academies Press. <https://doi.org/10.17226/10019>.

Pellegrino, James W., and Edys S. Quellmalz. 2010. “Perspectives on the Integration of Technology and Assessment.” *Journal of Research on Technology in Education* 43 (2): 119–34. <https://doi.org/ggfh8z>.

Pereira, Diana, Maria Assunção Flores, and Alexandra Barros. 2017. “Perceptions of Portuguese Undergraduate Students about Assessment: A Study in Five Public Universities.” *Educational Studies* 43 (4): 442–63. <https://doi.org/gmb98n>.

Pereira, Diana, Irene Cadime, Gavin Brown, and Maria Assunção Flores. 2021. “How Do Undergraduates Perceive the Use of Assessment? A Study in Higher Education.” *European Journal of Higher Education*, January, 1–17. <https://doi.org/gk36vq>.

Plake, Barbara S. 1993. “Teacher Assessment Literacy: Teachers’ Competencies in the Educational Assessment of Students.” *Mid-Western Educational Researcher* 6 (1): 21–27. <http://www.mwera.org/MWER/documents/MWER-1993-Winter-6-1.pdf>.

Plake, Barbara S., James C. Impara, and Jennifer J. Fager. 2005. “Assessment Competencies of Teachers: A National Survey.” *Educational Measurement: Issues and Practice* 12 (4): 10–12. <https://doi.org/fw7g9p>.

Popham, W. James. 2011. “Assessment Literacy Overlooked: A Teacher Educator’s Confession.” *The Teacher Educator* 46 (4): 265–73. <https://doi.org/dctz5h>.

Pressey, S. L. 1927. “A Machine for Automatic Teaching of Drill Material.” *School & Society* 25: 549–52.

Price, Margaret, Jude Carroll, Berry O’Donovan, and Chris Rust. 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): 479–92. <https://doi.org/d4sz5m>.

Renckly, Thomas R. 2002. *Air University Sampling and Surveying Handbook*. United States of America Department of Defense.

Schneider, Christoph, Christopher DeLuca, Marcela Pozas, and Andrew Coombs. 2020. “Linking Personality to Teachers’ Literacy in Classroom Assessment: A Cross-Cultural Study.” *Educational Research and Evaluation* 26 (1-2): 53–74. <https://doi.org/gk5r6r>.

Scriven, Michael. 1967. “[The Methodology of Evaluation](https://www.ncbi.nlm.nih.gov/pmc/articles/null).” In *Perspectives of Curriculum Evaluation*. Rand McNally Education Series. Rand McNally.

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

Smith Glasgow, Mary Ellen, H. Michael Dreher, and James Schreiber. 2019. “Standardized Testing in Nursing Education: Preparing Students for NCLEX-RN® and Practice.” *Journal of Professional Nursing* 35 (6): 440–46. <https://doi.org/ggqh9p>.

Smith, Lisa F., Mary F. Hill, Bronwen Cowie, and Alison Gilmore. 2014. “Preparing Teachers to Use the Enabling Power of Assessment.” In *Designing Assessment for Quality Learning*, edited by Claire Wyatt-Smith, Valentina Klenowski, and Peta Colbert, 1:303–23. The Enabling Power of Assessment. Dordrecht: Springer Netherlands. <https://doi.org/10.1007/978-94-007-5902-2_19>.

Solomonidou, Georgia, and Michalis Michaelides. 2017. “Students’ Conceptions of Assessment Purposes in a Low Stakes Secondary-School Context: A Mixed Methodology Approach.” *Studies in Educational Evaluation* 52 (March): 35–41. <https://doi.org/f9w8gd>.

*Standards for Educational and Psychological Testing*. 2014. Washington, D.C: American Educational Research Association. <https://www.testingstandards.net/uploads/7/6/6/4/76643089/9780935302356.pdf>.

Stiggins, Richard J. 1991. “Assessment Literacy.” *The Phi Delta Kappan* 72 (7): 534–39. <http://www.jstor.org.ezproxy.library.uvic.ca/stable/20404455>.

———. 1995. “Assessment Literacy for the 21st Century.” *The Phi Delta Kappan* 77 (3): 238–45. <http://www.jstor.org.ezproxy.library.uvic.ca/stable/20405538>.

Tashakkori, Abbas, Burke Johnson, and Charles Teddlie. 2020. *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Second Edition. Thousand Oaks: SAGE Publications, Inc.

Tekir, Serpil. 2021. “Alignment of the Intended, Enacted, Received and Assessed Curriculum in EFL Pre-Service Measurement and Evaluation Education.” *Education and Science*. <https://doi.org/gmgbpz>.

Timmis, Sue, Patricia Broadfoot, Rosamund Sutherland, and Alison Oldfield. 2016. “Rethinking Assessment in a Digital Age: Opportunities, Challenges and Risks.” *British Educational Research Journal* 42 (3): 454–76. <https://doi.org/gftz95>.

Watson, Cate, Anna Wilson, Valerie Drew, and Terrie Lynn Thompson. 2017. “Small Data, Online Learning and Assessment Practices in Higher Education: A Case Study of Failure?” *Assessment & Evaluation in Higher Education* 42 (7): 1030–45. <https://doi.org/ggrgxn>.

Watters, A. 2021. *Teaching Machines: The History of Personalized Learning*. MIT Press. <https://books.google.ca/books?id=zukGEAAAQBAJ>.

Watters, Audrey. 2015. “Speak & Spell: A History.” Hack Education. January 13, 2015. <http://hackeducation.com/2015/01/13/speak-and-spell>.

Webb, Mary, and Dirk Ifenthaler. 2018. “Assessment as, for, and of Twenty-First Century Learning Using Information Technology: An Overview.” In *Second Handbook of Information Technology in Primary and Secondary Education*, edited by Joke Voogt, Gerald Knezek, Rhonda Christensen, and Kwok-Wing Lai, 581–600. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-71054-9_37>.

Willis, Jill, Lenore Adie, and Val Klenowski. 2013. “Conceptualising Teachers’ Assessment Literacies in an Era of Curriculum and Assessment Reform.” *The Australian Educational Researcher* 40 (2): 241–56. <https://doi.org/gh5k7d>.

Woldeab, Daniel, and Thomas Brothen. 2019. “21st Century Assessment: Online Proctoring, Test Anxiety, and Student Performance.” *International Journal of E-Learning & Distance Education* 34 (1). <https://www.ijede.ca/index.php/jde/article/view/1106>.

Xu, Yueting, and Gavin T. L. Brown. 2016. “Teacher Assessment Literacy in Practice: A Reconceptualization.” *Teaching and Teacher Education* 58: 149–62. <https://doi.org/f8txgm>.