

Towards a Situated View of Assessment Literacy for Higher Education

A Thesis Submitted to the College of Graduate and Postdoctoral Studies in Partial Fulfillment of
the Requirements for the Degree of Master of Education in the Department of Educational
Psychology and Special Education

University of Saskatchewan

By

Derek Friesen

Permission to Use

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other uses of materials in this thesis in whole or part should be addressed to:

Head of the Department of Educational Psychology and Special Education
28 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan, S7N 0X1
Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
105 Administration Place
Saskatoon, Saskatchewan, S7N 5A2
Canada

Abstract

The guiding purpose of this study was to explore how the term *assessment literacy* (AL) could be differently constructed in higher education (HE) settings as opposed to how it has been constructed for other settings. First, an empirical scoping review of 182 sources revealed *AL for HE* as more sophisticated than described with current AL models. Emergent themes of the scoping review were compared to existing AL conceptualizations and discussed with consideration to prior AL research. Using the scoping review results as a theoretical framework, a survey was developed to investigate how common assessment tasks in HE settings could be organized and labelled, and how such tasks may differently invoke speculated components of AL for HE. The thesis reports on survey development, item pool review, and various statistical analyses of data from a limited convenience sample of faculty from a Western Canadian HE institution. Survey findings revealed that assessment is implicated in a range of HE tasks that seem separable and which have not been considered in previous literature regarding AL. Further, different HE tasks were associated with varying strength to different arrangements of components of AL. As well, components of AL that were speculated from thematic analysis of a body of literature related to AL in HE were correlated to one another in various ways. Together, these findings indicated that AL may be explained by examining composite sub-concepts in relation to one another, and further that such a model for AL is related to different tasks in different ways. This underscores the idea that particular purposes for assessment in HE may be nested within a more general conceptualization of AL. Limitations and directions for future work on conceptualizing AL for HE are also discussed.

Keywords: *assessment literacy, higher education, scoping review, survey study, mixed methods, factor analysis, thematic analysis*

Acknowledgements

I am grateful to Dr. Amin Mousavi who supervised, guided, provided opportunities to grow and continually reassured throughout this effort. Thank you for your patience, prompt feedback, and the constant refrain to seek functional solutions instead of complicated ones. Thank you for the opportunity to take on this project and for reminding me that practicality and attention to detail can coexist in the same universe.

Thank you to committee members Dr. Hellsten and Dr. Newton who provided valuable feedback and oversaw a defense with relatively short notice. I appreciate your willingness to participate and your flexibility.

Thank you to the assessment experts who participated in an item review. Throughout this project I have become intimately familiar with much of your work. Your names and citations will be forever etched into my mind.

Thank you to Kaitlyn Dueck, my dear partner. I am inspired daily by your effort, your advocacy, and your thirst for possibility. Thank you for your patience with my process and your keen discernment about rest and balance.

A final thank you to a legion of friends, colleagues, and coworkers who listened politely and nodded at all the right times as I tried to explain my efforts.

I would also like to un-acknowledge the novel coronavirus and all of its wretched descendants. I know that microbial life is dominant on this planet, but this whole pandemic affair has been patently unreasonable, and I do not care for it.

Table of Contents

Towards a Situated View of Assessment Literacy for Higher Education	i
Permission to Use	ii
Abstract.....	iii
Acknowledgements.....	iv
Table of Contents.....	v
CHAPTER ONE: THESIS OVERVIEW	1
1.1 <i>Significance and Purpose</i>	1
1.2 <i>Research Questions and the Present Study</i>	2
1.3 <i>Parameters</i>	3
1.3.1 Assumptions.....	3
1.3.2 Delimitations	4
1.4 <i>Definitions</i>	5
1.5 <i>The Researcher</i>	6
1.6 <i>Thesis Organization</i>	6
CHAPTER TWO: LITERATURE REVIEW	8
2.1 <i>Assessment in Higher Education</i>	8
2.1.1 Measurement, Assessment, and Evaluation	10
2.1.2 Unique Features of Higher Education Assessment.....	13
2.2 <i>Emergence of Assessment Literacy</i>	14
2.2.1 Conceptual History of Assessment Literacy	14
2.2.2 Assessment Literacy is Situated in Context.....	16
2.3 <i>Assessment Literacy in Higher Education</i>	17
2.3.1 Lessons from Language Assessment Literacy	20
2.4 Theoretical Approach for a Conceptualization of AL For HE.....	21
CHAPTER THREE: SCOPING REVIEW METHODOLOGY	23
3. <i>Search Strategy and Results</i>	23
CHAPTER FOUR: SCOPING REVIEW ANALYSIS	28
4.1 <i>Thematic Analysis</i>	28
4.2 <i>Research Question One</i>	30
4.2.1 Conceptualizing AL in HE.....	30
4.2.2 Knowledge and Practice for HE Assessment.....	33
4.2.3 Assessment Fluency	34
4.2.4 Metacognition	34
4.2.5 Conceptions of Assessment.....	35
4.2.6 Context	35
4.2.7 Decision-making	36
4.2.8 Development	36
4.3 <i>Research Question Two</i>	37

4.3.1 Quality Assurance Processes and External Evaluation.....	38
4.3.2 AL for Academic Developers	39
4.3.3 Language Assessment Literacy.....	39
4.3.4 Professional Roles and Assessment Literacy	42
4.4 <i>Scoping Review Summary and Theoretical Framework</i>	42
CHAPTER FIVE: SURVEY METHODOLOGY	44
5. <i>Survey Development</i>	44
5.1 Development of the Item Pool	45
5.2 Broad Appeal and Item Review	48
5.3 Reconceptualization and Redevelopment	50
5.4 Deployment Strategy.....	55
5.5 Data Collection.....	55
5.6 Plan for Statistical Analysis	56
CHAPTER SIX: SURVEY RESULTS AND ANALYSIS	57
6.1 <i>Sample Description</i>	57
6.2 <i>Research Question One</i>	59
6.3 <i>Research Question Two</i>	62
6.4 <i>Research Question Three</i>	68
6.4.1 Relationships Between HE Assessment Tasks and AL Components	72
6.5 <i>Research Question Four</i>	72
6.5.1 Relationships Among AL Components	73
6.5.2 Independent t Tests	73
6.6 <i>Thematic Analysis of Qualitative Items</i>	76
6.6.2 Open Ended Item One	76
6.6.2 Open Ended Item Two	79
6.7 <i>Summary of Survey Findings</i>	81
CHAPTER SEVEN: DISCUSSION AND CONCLUSION	83
7.1 <i>Research Question One</i>	84
7.2 <i>Research Question Two</i>	86
7.3 <i>Research Question Three</i>	87
7.4 <i>Research Question Four</i>	89
7.5 <i>Discussion Summary</i>	91
7.6 <i>Limitations</i>	92
7.7 <i>Future Directions</i>	95
References.....	98
Appendix A: Thematic Summary	113
Appendix B: Themes of Assessment Professionalism	126
Appendix C: Certificate of Ethical Approval	129
Appendix D: Survey Items	130

Appendix E: Frequency of Assessment in HE Tasks Correlation Table	138
Appendix F: Item Review Recruitment Materials and Consent Form	139
Appendix G: Survey Consent Form	142

CHAPTER ONE: THESIS OVERVIEW

Effective faculty involvement in assessment in higher education (HE) involves clear comprehension of the theory and principles that underlie educational assessment to such a degree that they may be appropriately implemented in practice, as well as discussed with other stakeholders in assessment processes (Davies & Taras, 2018; Medland, 2019; Popham, 2011). Such effective application has been referred to as *assessment literacy* (AL) which is increasingly implicated as vital for professional faculty in HE settings (DeLuca et al., 2019; Medland, 2018; Willis, Adie, & Klenowski, 2013). Contemporary views of AL assert the prominence of contextually situated skills, “views, and knowledge that are integral to confidence and efficacy for implementing assessment-related initiatives” (Fulmer, Lee, & Tan, 2015). In this way, AL is viewed as a profile of competencies necessary for individuals to assess effectively in particular settings and for specific purposes.

Previous conceptual and empirical work has primarily explored AL as pertaining to how K-12 teachers assess student learning (e.g., Stiggins, 1995; DeLuca, LaPointe-McEwan, & Luhanga, 2015). The concept of AL has since been utilized to describe necessary competencies for assessing student learning in HE settings (e.g., Davies & Taras, 2018), as well as specific content domains such as language assessment (e.g., Taylor, 2013; Kremmel & Harding, 2020). Given the widespread use of assessment for academic and nonacademic decision-making, it is pertinent to conceptualize and promote AL among stakeholders who conduct assessment activities in HE contexts. At present, AL for HE has yet to be sufficiently differentiated from AL as conceptualized for other contexts. Specifically, there is a lack of understanding about how assessment is used by HE faculty in everyday professional activities for purposes beyond assessing student learning, such as determining faculty advancement or participating in departmental initiatives (e.g., Arreola, Theall, and Aleamoni, 2003). To represent all facets of how faculty actively assess in their everyday work it was pertinent to explore how AL may be differently defined for HE, focusing on how assessment is involved in all domains of HE professional work.

1.1 Significance and Purpose

This study was important because assessment seemed to be implicated in many HE decision-making functions including advancement, recruiting, student development, and research. Prominent AL conceptualizations seemed to center on pedagogical practices of assessment. Investigations have predominantly focused on how teachers in school-based education understand assessment concepts relevant to their role (e.g., formative and summative assessment) and relate them to one another (DeLuca et al., 2015; Xu & Brown, 2016). However, there seemed to be a dearth of specific conceptual development of AL as uniquely situated in HE contexts. Because of this, the construct of AL as it has been understood to date appeared incongruent with the assessment processes involved in HE settings. As such, it seemed necessary to explore how faculty and professionals in HE contexts understand and engage in assessment

activities in different ways than in other educational contexts. To this end, a scoping review summarized how AL has been conceptualized for HE contexts, and HE faculty were surveyed to investigate how they perceive assessment as involved in their everyday work.

The guiding purpose of this thesis was thus to explore how the term AL could be differently constructed in HE settings. This was useful for understanding AL as a profile of competencies that are utilized in different constellations by faculty in different roles. Understanding how faculty in different roles use particular components of AL allows for a more targeted and modular approach to faculty development in HE. There were five primary objectives to this thesis. First, to develop a comprehensive understanding of HE assessment by determining encompassing elements, features, procedures, and actions that have been implicated in academic research. Second, to establish a thematic framework synthesizing research literature on AL in HE. Third, to establish preliminary empirical support for AL in HE as conceptualized by research literature. Fourth, to develop guidelines supporting HE administrators and academic developers in design and delivery of professional development programs in AL. Fifth, to disseminate research findings and mobilize knowledge of AL in HE to education community contexts.

1.2 Research Questions and the Present Study

The research problem targeted by this thesis was the lack of an AL representing a wholistic picture of the uses and purposes of assessment in HE settings. To begin to address this problem, a multifaceted conceptualization of AL for HE seemed viable to scaffold research and professional dialogue regarding AL beyond the historical focus on student learning assessment-related activities. Such a conceptualization could not be bound to any one specific context in HE; rather, it must act as a ‘higher order’ concept in which more specific literacies (e.g., Language AL or student learning AL) can be situated. Thus, the purpose of first phase of this study was to synthesize academic literature on the terms “assessment literacy” and “higher education” to delimit a body of academic work distinguishable from how the term AL has been used in school-based settings, but which also accounts for the broad variety of purposes and professional responsibilities assumed by faculty in HE. To this end, guiding research questions for the scoping review were:

- 1) What are the elements and features of “assessment literacy” in higher education that set it aside from school-based (pre-K through 12) education? And,
- 2) How can these elements and features be integrated in a conceptual model to develop a useful unified definition of assessment literacy for higher education settings?

To being investigating these questions, this thesis used scoping review methodology to broadly map existing academic literature regarding AL in HE. The scoping review proceeded according to an established procedure for such a project (Peters et al., 2015): (1) identifying the research question; (2) identifying relevant studies; (3) selecting studies; (4) charting studies; (5)

collating and summarizing; and (6) reporting results. A scoping review was appropriate given the non-specificity with which AL appeared to have been framed in HE contexts, as opposed to a systematic review in which more specific criteria may exclude relevant research to AL for HE. Following the review, a general inductive approach (i.e., Thomas, 2006; Mykota, & Muhajarine, 2005; Hootz, Mykota, & Fauchoux, 2016) was used to determine the themes and their respective subthemes emerging from the studies.

The scoping review resulted in the identification of gaps in literature related to the non-student assessment tasks performed by higher education faculty. It was found that the concept of AL as applied to HE settings tended to focus on the competencies needed for processes involved in assessing students, such as classroom assessment, SLO/SBA, formative/summative assessment, student assessment design, conceptions of student assessment, or decision-making using student assessment data. The review revealed an impetus to establish an understanding of the professional tasks of HE faculty as related to assessment literacy. To accomplish this, a survey study investigated the following research questions with an exploratory mixed-methods survey of HE faculty in a Canadian prairie university:

- 1) What tasks are HE faculty performing in their everyday professional contexts that involve assessment?,
- 2) How can common HE assessment tasks be organized and labelled?,
- 3) What relationships exist between common HE professional tasks and theorized components of AL for HE? And,
- 4) How is HE assessment and its involvement in HE professional work uniquely construed by HE faculty?

1.3 Parameters

The parameters of the study include its assumptions, delimitations, and definitions of several key terms. Assumptions are described to contextualize the project ontologically, whereas delimitations describe reasonable boundaries to the interpretation of the findings. Using these parameters, the following sections describe the context and conditions in which the present research occurred.

1.3.1 Assumptions

The methodology of the study was a practical mixed-methods design occurring in two parts. In the first phase, a scoping review was conducted to synthesize themes from recent research related to AL in HE. It was assumed that a scoping review would result in a thematic synthesis of how the term AL has been used pertaining to HE. The results of the scoping review study were girded by several assumptions. First, it was assumed that available academic literature represented elements, features, and aspects of HE assessment that may be considered essential for effective assessment practice in HE. Next, it was assumed that utilizing standardized scoping review methodologies would be sufficient to accurately elicit such elements, features,

and aspects of HE assessment from a large body of conceptually incongruent academic research. Third, it was assumed that thematic analysis of such literature would accurately identify aspects of AL unique to HE contexts.

In the second phase, these themes informed development of a questionnaire to survey HE faculty about how they perceive components of AL and involve assessment in their work. This phase used a survey design by developing a survey to gather numerical and qualitative (i.e., open-ended questions) data on how HE faculty involve assessment in their work and data regarding speculated elements of their assessment literacy. These data were analyzed using quantitative methods including exploratory factor analysis to explore the underlying dimensionality of the survey responses and correlation analyses to explore relationships between speculated factors and components of AL. It was assumed that choosing such a design would best inform a useful conceptualization of AL for HE by facilitating both researcher and practitioner perspectives from the bottom up. The results of the survey study were girded by the following assumptions. First, that the questionnaire was developed in a way that reflected sound practices for scale development to develop an instrument to elicit exploratory data in valid and reliable ways. Next, it was assumed that participants responding to the survey did so voluntarily, providing informed consent and responding to the questionnaire in ways that reflected their genuine experience with assessment in HE. Finally, it was assumed that no errors were made during data collection, entry, and analysis. Statistical analyses were selected to be as straightforward and practical as possible, and interpretations of the analyses were made as conservatively as the limited exploratory data allowed.

The thesis project itself is underwritten by an assumption that *assessment* is a general process of collecting and documenting data for decision-making which is applied to specific contexts such as student learning or program effectiveness. Different aspects of assessment are important in different situations dependent on context as espoused by Willis, Adie, and Klenowski (2013). To this point, the thesis assumes a subject-oriented theoretical framework in which assessment and AL are framed through the lens of HE faculty who use it for different reasons in varying amounts in their everyday professional work (Jurczyk, Vob, & Weihrich, 2016). In this way, the thesis is approached with an emphasis on how HE faculty conduct assessment activities within the context of their everyday life, and how meaning-making is entwined with the context of the assessor (Schraube & Marvakis, 2016). Such a subject-oriented frame is comparable to contemporary pragmatist epistemologies such as that espoused by Rorty (2009), which reject a priori theorization in favour of the view that truth is tied to the practical living of life.

1.3.2 Delimitations

The scoping review was delimited to contemporary research targeting AL for HE faculty including lecturers, researchers, development staff, and administrative stakeholders. Included studies were published in English including empirical studies, survey studies, advocacy literature, systematic reviews, meta-analyses, and thesis and dissertation papers. Given the wide

net cast by a scoping review, sources were included that appeared to relate to AL without explicit statements, including terms like “assessment knowledge” or research articles looking at the evaluation/evaluative skills of faculty members, or how faculty beliefs influence assessment understanding. This determination was made in an effort to over-reach rather than under-represent potential facets of AL for HE.

The survey study was delimited by the setting and context of the gathered data. Survey data were gathered from HE faculty in the Prairie Provinces regardless of gender, length of service, area of study, or position held. As such, generalizability of interpretations are not generalizable to wider HE institutional contexts either within Canada or internationally. The survey sample was acquired for its convenience in multiple ways. First, recruiting participants from the local institutional context required only the ethical approval from the local ethical review board, whereas sampling from a wider swath of universities would have required ethical approvals from each given institution. Next, HE faculty are considered a difficult population to elicit response from, especially amid additional workload and adjustments to everyday professional life due to the COVID-19 pandemic. A major delimitation was the inability to achieve a targeted sample size due to response attrition, time restraints, and cost restraints. Despite these boundaries, the results of the study were expected to have interesting applications to HE practice outside of Prairie University contexts, and ultimately to serve as a starting point for further investigations into the AL of HE faculty.

1.4 Definitions

Assessment: A process of garnering and documenting specific empirical data about *specific* phenomena in relation to a purpose or goal (Kizlik, 2012; Overton, 2012). It *may* use measurement “as a special kind of evidence” but may also use non-quantitative sources of data (Ewell & Cumming, 2017). Comparing *assessment* to *measurement*, Cumming and Ewell (2017) highlight how *The Standards for Educational and Psychological Testing* (AERA, APA & NCME, 2014) delineate how assessment refers to a process broader than measurement “that integrates [individual test] information with information from other sources” (p. 11) as opposed to the specific technical procedures and principles composing educational measurement (Overton, 2012). Further elucidation of how *assessment* has been used in previous literature and is used here in present in Chapter 2.

Assessment Literacy (AL): A collection of related and contextually situated competencies necessary for implementing assessment-related initiatives (DeLuca et al., 2018; Fulmer, Lee, & Tan, 2015; Willis, Adie, & Klenowski, 2013). In higher education, Davies and Taras (2018) describe how “*assessment literacy*, in any given setting, may be defined as an understanding of the issues, general and specific criteria, and standards which may enable a given individual to communicate efficiently with individuals in a similar context and also to negotiate meaning (coherently) from an informed position, on assessments of processes or products made within that context” (p. 475-6).

Higher Educational Assessment: The process of gathering, documenting, analyzing, and interpreting empirical evidence to make decisions related to situations like (but not limited to) assessing student learning, hiring/advancement, program evaluation, admission, and promotion.

1.5 The Researcher

The principal researcher for the thesis resides in the setting and context where the research was conducted. The researcher holds a Bachelor of Arts degree in Behavioural Science, and the present thesis was conducted for the purposes of fulfilling the requirements for completing a master's program in Educational Psychology specializing in Measurement and Evaluation. The researcher has an interdisciplinary background, having previously conducted qualitative research investigating emerging adults' experiences of anxiety and engaging in further academic work in the fields of critical and cultural psychology. The researcher is an active paraprofessional as a psychometrist in a developmental pediatric health care setting, administering standardized psychoeducational assessments to support complex developmental and psychological diagnoses.

1.6 Thesis Organization

The thesis is organized in a nontraditional format to reflect the chronological process of the study. The following chapters include a literature review (Chapter Two), then a section describing methodology for the scoping review study (Chapter Three), and the results and analysis from the scoping review (Chapter Four). For the second phase of the thesis, the survey methodology is described (Chapter Five) followed by the results and statistical analyses of the survey data (Chapter Six). The final chapter is a discussion of the implications of the findings (Chapter Seven). The literature review offers contextual background information about educational assessment and AL, as well as summaries of previous research that justify the necessity of the present study. The importance of the study is outlined by summarizing educational assessment in Canada, which is followed by a discussion of the development of AL and implications for higher education settings. For the scoping review methodology, a linear description of the review is provided including methodology, procedures, results, and thematic analyses of scoping review data. Following this, A results chapter reports and describes the scoping review themes and how they establish a theoretical framework for the survey study. The survey methodology chapter describes the survey study with a description of the questionnaire that was developed to investigate AL, the participants that were recruited, the step-by-step procedures involved in the study, and a description of the statistical analyses that were used. The survey results chapter reports and describes the scoping review themes and how they informed the survey study. It is followed by analysis of survey data from statistical and thematic analyses with pertinent information tabulated. The closing section, the discussion, interprets the research findings, and explores their implications in terms of the current literature and for practice and future research.

CHAPTER TWO: LITERATURE REVIEW

This chapter summarizes literature on assessment and AL in HE to exemplify gaps pertaining to purposes other than assessing student learning. A brief exploration of assessment in educational contexts is provided, contrasting it with oft-confused terms *measurement* and *evaluation* as well as delimiting a common educational overspecialization of the term to mean *student learning assessment*. The context in which AL has been developed is discussed presenting different contemporary definitions and approaches to AL and potential corresponding components. Just as assessment is more than assessing learning, AL must involve knowledge, competency, and fluency in assessment for a diverse range of profession-specific purposes and tasks. This is connected to HE by reviewing assessment-based activities in HE: the review demonstrated a dearth of research on AL in HE, and a particular paucity of research examining AL for non-student learning assessment purposes in HE. This sets the basis for a two-phased project involving a scoping review and a survey study which are discussed in the following chapter. Note that more comprehensive critique of AL in HE literature may be found in Chapter Four, which integrates such critique with the scoping review analysis.

2.1 Assessment in Higher Education

Assessment is a component of everyday academic life taking different forms and requiring various competencies. It is an onerous, time-consuming, and yet critical aspect of any educational context (Huber & Skedsmo, 2016). The emergence of assessment as a professional domain of scholarship and expertise has seen the term *assessment* evolve in meaning and usage. Reviewing assessment as an emergent domain of semi-professional practice, Ewell and Cumming (2017) helpfully explored the history of how the term *assessment* has been classically used in different ways in educational contexts. One critical observation from their review is how different educational traditions came together in the mid 1980s to espouse how “colleges and universities as institutions could “learn” from feedback on their own performances and that appropriate research tools were now available for them to do so” (p. 8). This reflection is important to the present argument in two ways. First, it highlights *assessment* as a processive act utilizing different strategies for different purposes, a convention contemporarily understood as a “contextual situation”. Second, it reveals how early assessment efforts were not dominantly directed towards *student learning assessment*, but rather a recognition of empirical strategies that could be reliably and repeatedly utilized to measure and adjust performance in a range of professional domains.

After establishing a theoretical groundwork for the emergence of assessment, Ewell and Cumming (2017) described how *assessment* as a term was utilized in different ways for different purposes. In the mastery-learning tradition, *assessment* was used to establish individual students’ mastery of complex abilities. Second, *assessment* referred to large-scale testing programs in K-12 contexts for accountability purposes through benchmarking school and division performance using psychometrically sound and efficient standardized examinations. Third, *assessment* was defined “as a special kind of program assessment” (p. 10) focused on using a range of mixed

method techniques for program and curricular improvement. More recently, *assessment* has been increasingly asserted as inseparable from teaching and learning processes (e.g., Brookhart, 2011). Reviewing these definitions it could be argued that the diverse ways that *assessment* has been defined is contemporarily problematic in terms of construct-specific usefulness. While Ewell and Cumming (2017) describe these different uses of assessment, they do not go as far as to account for how these definitions are used contemporarily in more specialized contexts. These purposes and functions of assessment are in use contemporarily, often without sufficient operationalization. The most contemporary understanding of *assessment* has not supplanted prior usage.

The convention of varied assessment definitions can be exemplified in Canadian HE contexts,. It has been argued that few generalizations can be drawn about educational assessment programs due to the decentralized nature of Canadian educational programming (e.g., Birenbaum et al., 2015; Volante & Jaafar, 2008). Each Canadian province and territory design their own policies for student assessment resulting in diversity in how various components of student learning assessment (e.g., summative, formative, alternative) are valued, and subsequently how fiscal and administrative resources can be directed towards these processes (Birenbaum et al., 2015). Despite the individuation of assessment programs among provinces and territories, some collaborative efforts have attempted to inform provincial assessment decision-making such as the *Principles for Fair Student Assessment Practices for Education in Canada* (Rogers, 1993). This document presents guidelines indicative of fair assessment as established by a committee comprised of representatives from professional associations and provincial/territorial representatives. Notably, this document swiftly abandons the specific term *student learning assessment* in favor of the brevity offered by merely using the term *assessment*. This underscores the same problematic convention within educational research wherein academics, educational professionals, and other stakeholders may frequently overlook sufficient operationalization of *assessment*, including specific purposes and tasks in context of use.

Conceptual meaningless due to overuse is not a new problem: consider early efforts in the field of intelligence testing, where Spearman (1927) famously claimed that “*intelligence* has become a mere vocal sound, a word with so many meanings that it finally has none” (p. 14). Hiser and Francis (2000) highlight how this problem has re-emerged throughout the study of intelligence, arguing that “there seems to be as many conceptualizations and definitions of intelligence as there are experts to write them” (p. 117). In a similar way, *assessment* seems to have become an overly general term used to refer to a range of more specific processes, creating problems where stakeholders operating under any given understandings of *assessment* may misinterpret other stakeholders’ usage. Thus, *assessment* has also become a word with such a volume of meanings and applications that, without context, it loses pragmatic value. In higher educational research, meaning can be reclaimed through the emerging popularity of specific terms like *program assessment*, *large-scale assessment* (LSA) or *assessment for learning* (AFL) which specifically describe the intended object of assessment efforts. Such linguistic specificity allows for targeted development of provincial/territorial assessment programs. Indeed, the past

several decades have seen Canadian educational systems respond to provincial accountability mandates by clarifying functions of student learning assessment. For example, *Rethinking Classroom Education with Purpose in Mind* (Earl & Katz, 2006) distinguishes subtleties among assessing *for*, *of*, and *as* learning to provide specific guidance for K-12 teachers who use assessments to monitor student learning and achievement. Earl, Volante, and DeLuca (2015) additionally specify AFL as more engaging, transparent, and inclusive than previous conceptualizations of formative assessment, and identify “an emerging effort across the country to integrate AFL” to enhance the student learning assessment practices of teachers.

In addition to using assessment techniques to monitor and support student learning, Canadian educational systems have experienced calls to accountability to prove that students receive the education they are promised (Brookhart, 2011; Volante & Jaafar, 2008). As a result there has been a proliferation of assessment techniques to either directly *improve* student learning (i.e., assessment for improvement) or *demonstrate* value to be gained through improvement (i.e., assessment for accountability), as well as communicate such evidence of improvement to the public at large (Ewell, 2009). Focusing the object of assessment as to produce evidence of student learning creates fertile ground for wider implementation of student learning outcomes (SLO) assessment which is viewed as a practical and time-efficient way for instructive staff to both improve classroom assessment and to demonstrate students have met learning goals on a course-by-course basis (Blair & Wise, 2011). Though this frame, other functions of educational assessment have been reframed in terms of their utility for improving student learning. For example, teacher effectiveness is assessed through the lens of an improvement/accountability dichotomy in terms of demonstrating added value through student improvement more so than by other components of teacher effectiveness such as innovation, resourcefulness, or participation in collaboration (Goe, Bell, & Little, 2008; Little, Goe, & Bell, 2009). In this way, assessment efforts not immediately oriented towards student learning become dichotomously characterized by utility for improvement of student learning or accountability to that purpose. The reductionistic tension between assessment for improvement and accountability, two common purposes for student learning assessment, thus problematically emphasizes a false dichotomy and centers debate around assessing student learning, eclipsing (and often demonizing) *assessment* for everyday professional decision-making purposes.

2.1.1 Measurement, Assessment, and Evaluation

It is only within the past half-century that *assessment* has factored so largely into educational language. Consider how Ewell and Cumming (2017) invoke the Latin origins of term assessment “ad + sedere” or “to sit beside”, as a symbol of a “process used to determine an individual’s mastery of complex abilities through observed performance” (p.10). They imply such *sitting with* to stem from direct observation for estimative purposes, but the etymological context “meant ‘to sit beside’ as would an assistant to the judge”, specifically for taxation purposes (Hodgson, 2008). Thus *assessment* has origins as a process to assist decision-making from a neutral standpoint rather than assign value through a power-over relationship, as a

‘master’ of learning holds over their student. *Assessment* as a term is contemporarily utilized as jargon-like vernacular in a myriad of unrelated contexts including financial assessment, property assessment, legal assessment, career assessment, or medical assessment. In each of these contexts, the term *assessment* describes a process of gathering and documenting empirical data for determination and decision-making in various capacities as needed for the specific professional context. Profession-specific usages of *assessment* thus comprise of minor semantic variations of this general process. For example, as opposed to a teacher an educational psychology professional referencing *assessment* likely refers to the use of a battery of psychoeducational tests, as well as observation and interview data, to gather evidence about student learning as well as social, behavioural, communicative, and adaptive functioning (Dombrowski, 2015). For a further example outside of educational contexts, assessment in a health treatment setting could imply a nurses’ competence in knowledge and practices of gathering information about a patients’ health and their ailments to determine health treatment needs and inform decision-making (e.g., Collins, Richmond & Buzila, 2018). On a smaller scale, the same nurse may conduct a ‘risk assessment’ before approaching any patient to make decisions about their own safe practice. It is clear that even within a given professional context, *assessment* is utilized in various forms for separable functions.

Clearly, the term *assessment* continues to be defined based on the epistemological needs of the profession or research at hand (Cumming and Ewell, 2017). Returning to educational contexts, the term *assessment* is often interchanged with terms such as *measurement*, *evaluation*, or *testing*. This has been problematic because these terms hold different meanings depending on the context of use. Consider how some academics use the terms *assessment* and *evaluation* interchangeably (e.g., Bastanfar, 2009), where others assert that they refer to entirely different processes (e.g., Kizlik, 2012). Still others argue that the terms reflect similar processes and methods directed toward different purposes (e.g., Apple & Krumseig, 1997). Given the differing use and definition by different stakeholders in the field of assessment, it is pertinent to distinguish among these terms and justify how this thesis uses the term.

Measurement refers to the act of quantifying attributes or dimensions – that is, discerning how something of interest can be represented numerically. In educational research, *measurements* are specific psychometric strategies used to estimate how much of a trait, attribute, or characteristic an individual possesses. From the established viewpoint of Classical Test Theory, a traditional comparison is drawn between measuring physical objects using standardized tools for measurement (e.g., rulers, scales, etc.) and measuring non-physical constructs of interests, which have no physical equivalent for standardization (i.e., the “amount” of a certain trait possessed by an individual (DeVellis, 2017; Kizlik, 2012). Ewell and Cumming (2017) explain how measurement may be confused with assessment “when describing the informational results of an assessment [...] implying that legitimate assessment should yield only quantitative results” (p. 25). Caution is necessary so as not to confound the quantifying practices of *measurement* with the broader processes of *assessment* or *evaluation*.

Evaluation, for that matter, implies processes of using collected and quantified information to make judgements and decisions for specific purposes in specific situations (Ewell and Cumming, 2017; Kizlik, 2012). *Evaluation* more closely approximates *assessment* in meaning than *measurement*. Etymologically, evaluation is concerned with *judging* value, worth, or quality (“Evaluation”, 2020), whereas assessment derives origins from *affixing* a value (“Assessment”, 2020). As described above, such affixing takes the position of neutral assistance to decision-making rather than the overall value judgment. A distinction can be drawn between objective and subjective processes underlying *affixing* and *judging*, respectively: *assessments* try to *objectively* collect data about something intangible to provide as much information as possible about a true and fair value, whereas *evaluations* use such information to make informed *judgements* involving such values. As such, some definitions of *evaluation* emphasize using gathered or available data to make *judgements* about the worth of educational activities (Academy of Process Educators, 2016; Apple & Krumsieg, 1997), compared to *assessment* which emphasizes a sense of *determination*, *ascertainment*, or *estimation*. In educational contexts, Cumming and Ewell (2017) describe evaluations as “evidence gathering processes that are designed to examine program or institution-level effectiveness” (p. 26), where ‘effectiveness’ is the standard by which worth is judged. However, other standards of ‘worth’ may also be evaluated in HE settings, such as cost-value or productivity (Wellman, 2013).

Returning to broad strokes, *assessment* describes a process of garnering and documenting specific empirical data about *specific* phenomena in relation to a purpose or goal (Kizlik, 2012; Overton, 2012). It *may* use measurement “as a special kind of evidence” but may also use non-quantitative sources of data (Cumming & Ewell, 2017). Comparing *assessment* to *measurement*, Cumming and Ewell (2017) highlight how *The Standards for Educational and Psychological Testing* (AERA, APA & NCME, 2014) delineate how assessment refers to a process broader than measurement “that integrates [individual test] information with information from other sources” (p. 11) as opposed to the specific technical procedures and principles composing educational measurement (Overton, 2012). *Evaluation* can be distinguished from *assessment* by a focus on judging worth of broad/aggregate processes in a particular context as opposed to the individual process focus of *assessment*.

Given such subtle definitional nuances, it is also pertinent to examine how the term *assessment* has been specified in educational settings. Most popularly, *student learning assessment* can be defined as an intuitive application of the general term *assessment* to the specific tasks and purposes associated with assessing student learning. A diverse body of research focuses on how assessment techniques can be used to promote growth, learning, or engagement among students (e.g., Black & Wiliam, 1998; Cauley & McMillan, 2010; DeLuca et al., 2018; Gilboy, Heinerichs, & Pazzaglia, 2015). However, it can be cumbersome to use these three words to repeatedly describe the same process. Thus, a recurring convention within educational research refers to *student learning assessment* as merely *assessment*. This contributes to linguistic confusion among researchers and practitioners who assume, and sometimes argue, that the general term *assessment* must refer exclusively to the specific processes of *student*

learning assessment. DeLuca et al. (2019) explain how generalization of the term *assessment* to mean *student learning assessment* may take origins from Bloom, Hastings and Madus's (1971) *Handbook on Formative and Summative Evaluation of Student Learning*, which classified *assessment* in relation to teaching considering its potential to enhance student learning and achievement. Another important influence was Black and Wiliams' (1998) extensive formative assessment review which popularly described classroom assessment-based teaching as a powerful tool with which teachers could target and scaffold student learning and emphasize life-long learning. Indeed, Cumming and Ewell (2017) argue the prime purpose of educational assessment as "to provide information that will enable faculty, administrators, and student affairs professionals to *increase student learning* by making changes in policies, curricula and other institutional programs" (p. 29). In doing so, they delimit the purpose of accountability as an important, but not driving, force for assessment. They argue that assessment *effort* ought to originate from academics and faculties interested in gathering and using data rather than from accountability mandates. While such ownership has indeed been found to bolster recurring assessment efforts (e.g., Schoepp & Tezcan-Unal, 2017) a more holistic perspective of educational assessment purposes and functions must not be overlooked.

2.1.2 Unique Features of Higher Education Assessment

In HE settings, Sarrico et al. (2010) argue that assessment use has multiplied due to changes in the macro-context surrounding HE such as massification, globalization, neoliberalization, expansion of private HE providers, and increasing competition. Further features of HE that impact assessment use include rapid expansion of higher education systems and modes of delivery; wider student participation; more diverse profile of institutions, disciplines, programs and students; rapid and growing integration of technology; internationalization; new and different financial models, as well as new and different models of governance (Tremblay, Lalancette, & Roseveare, 2012). In HE settings, faculty themselves participate in assessment of courses, programs, processes, and research products, as well as assessment of student learning (Arreola, Theall, & Aleamoni, 2003; Davies & Taras, 2018). These assessment tasks for which faculty may be responsible in HE settings may involve greater complexity or different nuance than assessment in other educational settings, such as that which occurs in K-12 education.

Despite how assessment may be defined in relation to HE, significant contextual variation exists regarding how assessment is used and understood among diverse HE faculties, colleges, and programs (Birenbaum, 2015). Some assessment-related activities in higher education may be more generalizable to faculty experience. These may include mandatory tasks like assessing student learning at the course level; participating in faculty evaluation processes using student, peer, or self-evaluation templates; or participating in program assessment to inform decisions regarding program development and curricular improvement. Such 'mandatory' assessment tasks in HE vary by department, program, college, and institution. Generalizations about what is involved in HE assessment must be made cautiously because of this contextual

variation, and as such one purpose of the present study is to examine elements of HE assessment that emerge as relatively common or ubiquitous to HE faculty professional expectations.

2.2 Emergence of Assessment Literacy

Much like the term *assessment*, AL has varied in meaning since its seminal use by Stiggins (1991), summarized by DeLuca et al. (2019) as an educators' "technical knowledge and skills in assessment, with a substantial emphasis on psychometric principles and test design" (np). Notably, such a definition maintains the assumption of assessment as predominantly oriented towards student learning without specifically invoking that purpose. Kremmel and Harding (2020) summarize how these early usages of assessment literacy "concentrated on primarily identifying components of assessment knowledge and skills of teachers" (p. 101).

Since this inception, the term AL has been subjected to broad use and nuanced definition by researchers with diverse epistemological backgrounds and purposes. Increasing emphasis has been placed on developing AL as a characteristic of educators at different levels, including how assessment is conceived of and standards for practice (Brown, 2004; Coombs et al., 2018; Hughes & Hargreaves, 2015). Some have attempted to characterize AL for measurement as a personal trait held by assessors (e.g., DeLuca, LaPointe-McEwan, & Luhanga, 2015). The concept of AL has widened to accommodate a popular focus on AFL (i.e., formative assessment) and the simultaneous falling out of summative student learning assessment (Brookhart et al., 2016). Such focus on AFL stresses the dynamic nature of classroom assessment, and as such AL has been framed in relation to how teachers approach assessment in response to everyday situations (DeLuca et al., 2019). As well, AL for teachers and school-based settings has been delineated in relation to other bodies of literature such as teacher education (Popham, 2011; Xu & Brown, 2016), including a knowledge base related to student and classroom assessment, conceptions of assessment, contextual boundaries, decision-making in practice, and eventual identity reconstruction to include the professional role of 'teacher-as-assessor'. Indeed, others have utilized AL as a trailhead to begin developing systemic comprehension of assessment among professionals (e.g., Medland, 2019; Taras, 2016). Such a wide range of conceptualizations and terminological usages emphasize how AL is increasingly referenced as contextually situated and dependent on response to a variety of situation-specific factors such as classroom context and personal background (e.g., Willis, Adie, & Klenowski, 2013, Looney et al., 2017).

2.2.1 Conceptual History of Assessment Literacy

As a construct, *assessment literacy* has been and continues to be used in diverse ways. Some use the term to clarify and strengthen existing assessment practices, (e.g., Davies & Taras, 2018), whereas other parties invoke AL to invite further future innovation in assessment (e.g., Popham, 2011). Gotch and French (2014) denote that it is generally agreed that AL emerged in the context of *The Standards for Teacher Competence in the Educational Assessment of Children* (AFT, NCME, & NEA 1990). Early efforts to understand AL sought to understand "the skills

and knowledge *teachers* require to measure and support student learning through assessment” (DeLuca, LaPointe-MacEwan, & Luhanga, 2015, p. 3). An instrument review conducted by Gotch and French (2014) described AL “as a stable but malleable attribute of teachers” (p. 14), identifying specific assessment behaviours such as using multiple high quality, targeted assessments, interpreting performance in the context of a particular assessment, appropriate administration and scoring, and communication of results. Here, it is seen how the early conceptual context implied AL as an inherently situated characteristic of teachers, rather than a construct that *can* be measured in teachers, but also in other professions. In school-based education, AL operates on the assumption that assessment efforts are primarily directed at evaluating student progress and enhancing student learning.

With this understanding in mind, DeLuca, LaPonte-MacEwan, and Luhanga (2015) identified how a range of efforts to measure teacher AL were consistently insufficient to represent a full range of student learning assessment activities. They denote how instruments developed to measure AL have utilized the *Standards* as a guiding framework, thus resulting in conceptual dismissal of over three decades of theoretical literature about the uses and purposes of educational assessment. DeLuca et al.’s (2015) review summarizes how Plake, Impara and Fager (1993) developed the Teacher Competencies Assessment Questionnaire (TCAQ), which identified gaps in teacher understanding of how to interpret, integrate, and understand assessment results. Further, several studies of pre-service and in-service teachers utilized the TCAQ alongside the Classroom Assessment Tasks survey (O’Sullivan and Johnson, 1993) and repeatedly identified deficits in assessment competency as conceptualized by the *Standards* (O’Sullivan & Johnson, 1993; Campbell, Murphy, & Holt, 2002; Mertler, 2003). Despite such deficiencies, DeLuca et al (2015) argue that the *Standards* themselves are insufficient frameworks to conceptualize assessment competency as they focus on summative uses of educational assessment. They explain how development of “Assessment for Learning” is an example of an important conceptual development that is overlooked in instruments attempting to measure teacher AL. This thesis agrees with this position, and furthers it by arguing that functions outside of student learning assessment are overlooked in most conceptualizations of AL.

Further conceptualizations of teacher AL have attempted to account for more contextual aspects of assessment. In a range of studies, Mertler and Campbell (2005) restructured the TCAQ as the Assessment Literacy Inventory (ALI), again identifying low teacher AL amongst preservice teachers. Efforts led by Brown (2004) and colleagues (e.g., Brown, Hui, Yu, & Kennedy, 2011; Brown & Remesal, 2012; Harris & Brown, 2009; Hirschfeld & Brown, 2009) to explore teachers’ conceptions of assessment (COA), eventually landing on a model that defines “teachers primary priorities related to the purposes of assessment with consideration for their values towards assessment practices” (p. 7). These efforts demonstrated growth in the understanding of AL as variable based on contextual factors. Finally, in response to Gotch and French’s (2014) review of teacher AL measures, DeLuca et al. (2015) developed the Approaches

to Classroom Assessment Inventory (ACAI), intentionally avoiding teacher AL terminology to create distance from the present/absent teacher AL dichotomy prevalent in previous literature.

Teacher AL was reconceptualized by Xu and Brown (2016) in relation to other bodies of school-focused literature, such as teacher education. It includes a knowledge base related to student learning and classroom assessment, conceptions of assessment, contextual boundaries, decision-making in practice, and eventual identity reconstruction as an assessor in that setting. The TALiP model asserts movement along a continuum of assessment development with three noteworthy levels: basic mastery, including knowledge of "what, why, and how" to assess, "without which teachers cannot engage with assessment at a deeper level" (p. 159). The second level involves integration of [student/classroom] assessment with principles of teaching and learning, graduating from an understanding of how assessment 'ought' to occur (in a best-practice sense) to a more personal conception that assessment, guided by sound theory, should be done. The highest level of mastery involves a "self-directed awareness of assessment processes and one's own identity as an assessor", implying metacognition and critical engagement with assessment principles and policies with contextual realities (for teachers, classroom realities). Elements of this highest level of mastery are summarized thus by the authors: "assessment literate teachers are those who constantly reflect on their assessment practice, participate in professional activities concerning assessment in communities, engage in professional conversations about assessment, self-interrogate their conceptions of assessment, and seek for resources to gain a renewed understanding of assessment and their own roles as assessors" (p. 159). This highest level of identity-as-assessor has features not necessarily specific to student assessment, which may also compose a broader AL that may be applied to specific professional foci in HE.

2.2.2 Assessment Literacy is Situated in Context

Recent conceptualizations of teacher AL are approached from a situated, contextual theoretical framework. This framework was initially espoused by Willis, Adie, and Klenowski (2013) who provide the most popular contemporary definition of teacher AL. They utilize a constructivist sociocultural frame where teacher AL is invoked to describe "dynamic social practices which are context dependent, and which involve teachers in articulating and negotiating classroom and cultural knowledges with one another and with learners" (p. 241). Here, AL is inseparable from the context in which it is situated, and thus the term and its components are defined by contextual construction rather than top-down application. Indeed, DeLuca et al. (2019) summarized more recent conceptions of teacher AL as a "differential and negotiated competency", emphasizing "the role of context in the capacity to develop and enact assessment knowledge and skills" (np). They list contextual factors affecting student learning assessment practice including teacher background, experience, professional learning, classroom contexts, student interactions, curriculum, and classroom diversity (Looney et al., 2017). Such factors impact how teachers enact student learning assessment practice across various professional domains such as understanding of assessment purposes, processes, conceptions of fairness, and

adherence to institutional or disciplinary standards. Ultimately, the recognition of AL as contextually situated provides important utility in facilitating the decoupling of AL from the context of teachers, and thus from a primary orientation towards student learning (e.g., DeLuca et al., 2018). It is thus reasonable that AL could also be situated in the context of higher educators, and differently so than other contexts. As evidenced in further sections, AL was recognized as a construct with descriptive utility in domains outside of teaching.

Given the range of tasks and purposes for which assessment is used, it is pertinent to approach AL from a pragmatist framework that champions variation in individual context from the ground up (e.g. Rorty, 2009). In an examination of contextual impacts on assessment, Fulmer, Lee, and Tan (2015) draw on systems approaches developed by Kozma (2003) and Bronfenbrenner (1977) to explain how prior research has focused on micro-contextual factors (e.g., personal/practitioner), has but neglected meso- (e.g., school/institutional) and macro-level (societal/cultural) explanations. Such a contextual theoretical approach is helpful to explain how AL has been inappropriately applied in HE as they help to compare the reality of everyday assessment work in HE to similar work in other lived contexts (e.g. Holzkamp, 2016). In a relevant framework, Fulmer, Lee, and Tan (2015) invoked “views and knowledge that are integral to confidence and efficacy for implementing assessment-related initiatives” (p. 1) to frame AL itself as a micro-level contextual influence mediating application of knowledge on implementation of assessment tasks. Here, AL is viewed as a component of assessment practice rather than an overarching concept suggesting assessment competence. In this way, any context where *assessment* is used may necessitate a contextually situated *assessment literacy*. AL is no longer bound to a specific profession, but rather a capacity that may be present (and thus, may be measured) in any context where assessment is used.

In summary, AL is contemporarily realized for teachers in school-based settings as a collection of contextually situated competencies “that are integral to confidence and efficacy for implementing assessment-related initiatives” (DeLuca et al., 2018; Fulmer, Lee, & Tan, 2015; Willis, Adie, & Klenowski, 2013). As such, teacher AL in school-based contexts can maintain the implicit assumption that assessment efforts are primarily directed at determining student progress towards learning outcomes and enhancing student learning. It is thought that teachers integrate various sources of knowledge with their own experiences to derive assessment practices relevant to their immediate context of use (Herppich et al., 2018). With a recognition of AL as contextually situated, these conventions can be directed towards AL in other settings indeed, the following section will describe AL in other such contexts.

2.3 Assessment Literacy in Higher Education

The diverse body of definitive work for AL was conducted prioritizing the experiences of teachers and learners’ school-based settings. As such, conceptualizations of AL in educational settings tend to retain the characteristic assumption of assessment being exclusively *as, of and for student learning*. As described, *student learning assessment* can be considered a specific educational application of *assessment* to the specific tasks and purposes associated with

assessing student learning. Clearly, some conceptual effort from the educational research community have resisted attempts to describe *assessment* as a professional competency beyond a dominant orientation towards student learning. In summarizing the history of higher educational assessment in the United States, Ewell and Cummings (2017) explained that “avoiding excessive professionalization [of assessment practice] was important because it promoted later linkages with the scholarship of teaching” (p. 18). This may be indicative of what Schoepp and Tezcan-Unal (2017) describe as a need for academic faculty to “see how their assessment reports and data [are] being used for course or program improvement” (p. 314) to perceive rationale for more broad approaches to HE assessment. This demonstrates how it is problematic to assume that assessment processes ought to be limited to student learning orientations. Though seemingly ubiquitous, student learning assessment does not account for every functional domain of assessment in HE. For example, Wellman (2013) exposed a dearth of literature examining a direct relationship between student learning assessment costs and outputs, highlighting how “student learning assessments have become too focused on compliance and do not yield results that are particularly helpful in making decisions about resources” (p. 17). This contention demonstrates the problematic nature of championing one purpose for assessment above others. The usefulness of assessment outcomes may be limited if one purpose (i.e., student learning assessment) is mandated as a priority at the cost of neglecting the applied value generated by other purposes. Instead, assessment efforts that generate pragmatic data within the everyday contexts of how faculty perform their daily work and pursue longer-term career goals.

Without undermining the functional importance of developing contextually situated skills and knowledge of student learning assessment, it is imperative to decouple the concept of AL from the student assessment practices implicated in school-based education settings. Faculty involvement of assessment in particular, contextually diverse HE settings requires professional expertise beyond the content of an average given academics’ field of study. Arreola, Theall, and Aleamoni (2003) espouse this idea by arguing that while HE faculty tend to be content experts in their given fields, they are also “expected to assume a variety of roles ... such as advising, serving on committees, or managing complex projects ... and perform at high professional level” (p. 1) in each of these undertakings. As such, Arreola, Theall, and Aleamoni describe the HE professoriate as a meta-profession wherein a faculty member requires the competencies of a base profession as well as the contextually combined elements “from a variety of several other professional arenas” (p. 2). For example, a faculty member in engineering must be competent in the subject-matter of engineering, but also require professional skills in teaching, assessing, researching, etc. Given these high demands, a generalized conceptualization of AL is critical for any faculty assuming professional responsibility for a variety of assessment-related tasks in a higher education context (Medland, 2019).

Some broader attempts to define AL outside of classroom and school-based settings have begun to generalize assessment efforts as not merely directed towards assessing student learning. For example, Davies and Taras (2018) described how “*assessment literacy*, in any given setting, may be defined as an understanding of the issues, general and specific criteria, and standards

which may enable a given individual to communicate efficiently with individuals in a similar context and also to negotiate meaning (coherently) from an informed position, on assessments of processes or products made within that context” (p. 475-6). Though Davies and Taras direct their efforts towards exploring faculty understanding of student learning assessment terminology, they began from a general frame of understanding assessment as it is used before moving into a specific application.

In a more general approach, Medland (2015, 2016, 2019) used inductive qualitative approaches in the context of external examination to explore general discursive components needed to achieve *shared understanding* in discussion and research: community, dialogue, knowledge and understanding, programme-wide approach, self-regulation, and standards. Medland intends for these elements to gird a shared discourse surrounding a conceptualization of AL for HE “that is fluid, applicable to different groups, and whose meaning can adapt to different contexts” (p. 577). Medland (2019) identified that AL tends towards deficit models in their description of AL as present (or high, or sophisticated) or, more likely, as absent (or low, or unsophisticated). Instead, Medland framed AL as a socially negotiated and personally developed variable rather than a binary and called for further work to understand AL in HE contexts.

Some specific professional domains have explored AL in relation to specialist competencies necessary to effectively make use of assessment of products and processes, such as LAL or assessment leadership. Examining AL in relation to these professional duties reveals multifaceted conceptualizations necessitating elements and features different to those necessary for school-based education. Conceptual components for assessment leadership in HE settings were explored through a modified Delphi process in HE pharmacy settings by Janke, Kelley, Sweet, and Kuba (2016). They grouped twelve competencies into three areas: context for assessment, managing processes of assessment, and leadership of assessment activities. These areas and competencies represent “roles for assessment professionals as experts, managers, and leaders” responsible for advancing assessment efforts and supporting departmental faculty as they develop their own AL (p. 6).

Regarding language assessment, Kremmel and Harding (2020) built on previous conceptual work (e.g., Baker, 2016; Taylor, 2013). They used exploratory factor analysis techniques on survey data to conceptualize *Language Assessment Literacy* (LAL) and extracted nine distinct components, some of which overlap with competencies for school-based AL, but many of which are specific to professional duties involved in HE language assessment. They further applied these competencies to profile different stakeholder groups in HE, enabling an understanding of the degree to which each competency is necessary for different professional purposes. Such an approach is relevant to the purpose of this thesis given the diverse professional needs for which assessment is invoked in general HE. The present project adopts a similar theoretical framework to that of Kremmel and Harding (2020); it is therefore pertinent to examine their conceptualization of LAL in more detail.

2.3.1 Language Assessment Literacy

In relation to the present project, *Language Assessment Literacy* (LAL) is a notable conceptualization of AL due to how it has been specified and developed for a particular context of professional work. LAL emerged in response to calls from the international community of language assessment professionals to explore professional AL as specifically pertaining to research, development, administration, interpretation, and implementation of data from language assessments (Taylor, 2013). Importantly, it is used and developed by language assessment researchers as a specialized form of AL existing and developing alongside but separate from a more general form of educational AL (Kremmel & Harding, 2020). In doing so, LAL recognizes and operationalizes assessment tasks specific to that domain of professional work. Some examples include developing language assessments, interpreting assessment results with psychometric rigor as useful within particular contexts, and using language assessment results in appropriate interprofessional ways to support academic decision-making such as qualification assessment and admissions decisions. This theoretical frame seems imperative and practical as it allows for definition of how assessments are used within a particular context while simultaneously recognizing that assessments are used differently in other contexts. This aligns with the subject-oriented frame of the present thesis by situating assessment within the context of assessment users as actors (e.g. Holzkamp, 2016). Indeed, LAL is recognized as necessary for diverse groups of social actors as stakeholders involved in using language assessments for decision-making processes. Kremmel and Harding note how early efforts to conceptualize LAL were directed at identifying potential “components of assessment knowledge and skills primarily required of teachers” (p. 102), but that LAL would be necessary for any stakeholders involved in designing, administering, and using data from language assessments including teachers, examination board personnel, language testing academics, and more (Taylor, 2009).

This leads to a further strength of LAL: conceptual resistance to labelling assessment stakeholders within a false dichotomy of being assessment literate or illiterate. Instead, LAL has consistently examined constituent *components* (Inbar-Lourie, 2008; Taylor, 2013) of AL along developmental continuums. Any given stakeholder is thus viewed as having capacity to increase (or decrease) their personal quantity of a particular component of LAL. It then follows that those professionals operating in unique contexts have specific levels of each LAL component, and as such groups of faculty in similar roles may have similar levels of each component as well. This line of thought led to a *developmental profiles* approach to LAL wherein specific professional roles possess quantities of LAL components as demanded by how they involve assessment with their everyday work (Taylor, 2013; Baker et al., 2014; Kremmel & Harding, 2020). Proponents of LAL have recognized that LAL is necessary for stakeholders both within the language assessment profession as well as interdisciplinary language assessment users. They agree that LAL is composed of separable components of assessment-related knowledge, skills, and beliefs that are particular to their context of use. Finally, professionals in different capacities likely view, need, and possess components of LAL in varying quantities that can be generalized as professional LAL profiles.

These assumptions about LAL are drawn from various theoretical frameworks. The idea that LAL is necessary is grounded in pragmatic views that a concept ought to resist *a priori* theorization, focusing instead on utility to the stakeholders at the basest levels and upwards (e.g., Rorty, 2009). Such a pragmatic orientation proliferates through educational settings, largely emphasizing the experience of the learner as the foci of educational endeavors and the role of instructors as directed toward student learning. Kremmel and Harding (2020) emphasize a focus on the everyday experience of practitioners who apply frameworks, recognizing that different stakeholders draw from concepts in different ways to achieve similar-yet-distinct goals. Somewhat conversely, the idea that practitioners possess quantities of components of LAL is based in frameworks championing psychometric measurement, such as Classical Test Theory (CTT). CTT maintains the distinction that individuals possess quantities of nonmaterial qualities that can be indirectly measured with sufficient attention to statistical and psychometric rigor. As such, LAL represents a blend of theoretical groundwork giving attention to both and the pragmatic experience of the individual as well as the standardization and validation practices of applied measurement.

These theoretical conventions are relevant to the purposes of the present study which intends to explore the elements and features of AL for HE contexts. Like language assessment, HE assessment involves a diverse range of stakeholders who are involved with assessment at different levels and for different functions, and as such likely involve assessment differently in their everyday professional work. Notwithstanding the assessment-related activities that may be more generalizable among faculty experience, HE faculty often elect to participate in faculty service opportunities based on their own contextual skills and interests, and for their own purposes. This may involve participation on boards for program-related purposes like student admissions; assessment and evaluation of thesis and dissertation papers; or evaluation of resource allocation for particular program goals and functions. Like LAL, it seems likely that HE faculty have different AL needs that correspond to how assessment is differently involved in their work based on contextual factors like discipline, seniority, or institutional assessment structures.

2.4 Theoretical Approach for a Conceptualization of AL For HE

In summary, isolated efforts have been made to examine components of AL in relation to specific professional disciplines, general academic discourse, and professional responsibilities in HE. The crucial step of acknowledging assessment as a process nonspecific to student learning seems to have been bypassed in HE AL literature, resulting in a too-narrow focus on literacy necessary for competent enactment of *student learning assessment* processes and negligence of broader HE assessment functions. For this project it is contended that higher education contexts involve assessment of products and processes beyond the setting and context of school-based education, and thus a conceptualization of *assessment literacy* accounting for elements and features specific to higher education is necessary for accurate definition. Such a

conceptualization enables continued research on the topic to be specific to and inclusive of the entirety of the higher education setting and relevant contextual influences.

Specifically, a sufficient view of AL in HE ought to be oriented pragmatically from the frame of faculty as assessors who use assessment against the backdrop of their everyday lives (e.g. Jurczyk, Vob, & Weihrich, 2016). A conceptualization of AL for HE ought to operationalize HE assessment in terms of how it is incorporated with professional work in HE outside of student assessment. Such a perspective recognizes *student learning assessment* as a distinct function and purpose of assessment particular knowledge and skills but give equal value to other functions and purposes of assessment common in the everyday work of faculty in HE. Next, a sufficient conceptualization of AL in HE will adopt a constructivist sociocultural frame as posited by Willis, Adie, and Klenowski (2013). These frameworks are increasingly invoked when conceptualizing AL for particular settings (e.g. DeLuca et al., 2019; Kremmel & Harding, 2020; Medland, 2018; Taras & Davies, 2018; Xu & Brown, 2016). This theoretical approach is appropriate for the present study as it acknowledges the wide contextual range of HE faculty including disciplinary background, institutional context, previous assessment experience, involvement in particular or additional faculty service roles. Grounding the present work in this framework responds to calls from other assessment researchers such as Medland (2018) who advocated for AL in HE as a socially negotiated and personally developed variable rather than a binary. For the purpose of the present study, contextual factors such as those espoused by Looney et al. (2017) are expected to inform individual faculty expression of AL in HE. Further to this, a sufficient conceptualization of AL for HE ought to establish separable components representing theoretical and applied assessment knowledge, skills, and beliefs necessary in varying amounts to particular kinds of HE work. This recognizes that faculty implement assessment in a multiplicity of roles beyond guiding student learning. In doing so, *assessment* itself is operationalized as a more neutral process of garnering and documenting specific empirical data about *specific* phenomena in relation to a purpose or goal (Kizlik, 2012; Overton, 2012), whereas AL situates assessment within a particular context and generate value in practice to those who use it. To this end, a sufficient conceptualization of AL for HE highlights fluency with assessment language to achieve shared discourse amongst HE professionals as a priority given the multiplicity of contexts in which assessment can be used.

To begin investigating a sufficient conceptualization of AL for HE, scoping review methodology was used to broadly map existing academic literature regarding assessment literacy in higher education. The scoping review project was conducted within the contextually situated framework enacted by similar approaches in other contexts. The following chapters describes the methodologies of both phases of the project beginning with the scoping review and concluding with the survey study that emerged in response.

CHAPTER THREE: SCOPING REVIEW METHODOLOGY

This chapter describes the methods used to address the research questions. A scoping review was conducted according to guidelines presented by Peters et al. (2015) to identify a comprehensive profile of literature relevant to AL for HE. The purpose was to outline the area of AL in HE and identify gaps in the existing body of literature to explore with future research. To identify a research question, a brief literature review broadly explored the topics of AL and assessment processes in HE settings. Notable influential studies included: Medland (2019, 2015), who developed discursive components for shared understanding of academic AL; Xu and Brown (2016), who conducted a scoping review exploring AL and teacher education; Fulmer, Lee, & Tan, (2015), who used systems approaches to describe contextual influences on teacher AL; and Taras and Davies (2013, 2017; Davies & Taras, 2016, 2018) who developed and implemented a questionnaire exploring various stakeholders discrimination of student learning assessment terminologies (i.e. summative and formative assessment). After reviewing literature the research questions were as follows: 1) What are the elements and features of “assessment literacy” in higher education that set it aside from school-based (pre-K through 12) education? And 2) How can these elements and features be integrated in a conceptual model to develop a unified definition of assessment literacy for higher education? To increase utility, feedback on these questions was received from a departmental scoping review team including university faculty, graduate and doctoral research assistants, and a university research librarian. This ensured the questions were as specific as possible and helped to delimit the scope of the review.

At the outset, the present scoping review is emphasized as exploratory in nature. As AL and its constituent components have not yet been situated in HE contexts, this review represents an initial effort to explore what components of AL for HE may exist and arrange them in a practical way. The body of research that emerged from this review should not be considered widely representative of a finalized, empirical account of AL for HE; instead, it ought to be framed as an exploration of the feasibility of situating AL within the context of HE.

3. Search Strategy and Results

To begin identification of relevant studies, a search strategy was developed with support from a university research librarian. It was determined to include studies published in the time span of January 1990 – December 2019, given that Stiggins first used the term AL in 1991. It was determined to include research targeting AL for HE faculty, including lecturers, researchers, development staff, and administrative stakeholders. Included studies would be published in English and comprise of research articles including empirical studies, survey studies, advocacy literature, systematic reviews, meta-analyses, and thesis and dissertation papers. Given the wide net cast by a scoping review, the search considered sources appearing to relate to AL without explicit statements. For example, this included sources using terms like “assessment knowledge”, research articles looking at the evaluation/evaluative skills of faculty members, or how faculty beliefs influence assessment understanding. These concepts were

surmised as related to AL due to their presence in the brief review of existing literature as denoted above. As per the purposes of the study to define AL specifically for HE faculty, it was determined to carefully exclude research conceptualizing the AL possessed by non-HE-faculty such as in-service/pre-service school-based (pre-k – grade 12) teachers, teacher candidates, school students, undergraduate students, and graduate students. The AL of HE students was of tangential interest, given that HE faculty may be responsible for developing their understanding of assessment. However, this project targeted the AL of HE faculty and it was important to ensure sufficient differentiation from existing conceptualizations for other populations. In general, sources targeting assessment elements, features, or skills and practices specific to non-higher education contexts were excluded. It ought to be emphasized that these exclusion criteria would alienate most of the literature explicitly dealing with AL, given the terms' origins and development in these contexts and settings.

In the first search phase, the University of Saskatchewan library search database was used to develop specific search strategy terms. The terms 'assessment literac*' and 'higher education' were used as subjects and keywords and combined using Boolean modifiers. An initial list of 169 relevant papers was compiled. To be considered relevant, titles and abstracts had to reference higher education as the target research context and to discuss concepts inferred to be related to assessment literacy. Subject headings and key words for each paper and their reference lists were listed as relevant to each concept (i.e., "assessment literacy" and "higher education"). As well, some keywords and subject headings were selected as specific exclusion terms to reduce the volume of results. For example, terms such as 'K-12', 'primary', 'elementary', and 'high school' were used to reduce results pertaining to non-higher educational contexts.

In the next phase three databases were selected for further search: ProQuest education, ERIC, and PsycInfo. These three databases were selected to triangulate sources published in different contexts, and to capture sources that were not accounted for by other databases. For example, searches in ProQuest Education returned a greater selection of theses and dissertations than the other two databases. Convenience was also a major factor in selecting these databases. These databases were readily recommended and available through the library at the home institution of the researcher. A large number of sources were selected, and these databases supported reference management functions that enabled efficient categorization, export, and identification of duplicates. However, there are sources of bias worth noting given the reliance of scoping review methodology on the use of academic data. One issue is publication bias: in several contexts, it has been shown that studies with significant results are more likely to be published (e.g., Egger & Smith, 1998; Pigott et al., 2013). Another issue is the omission of reporting nonsignificant results; Pigott et al. (2013) found that nonsignificant results were 30% more likely to be omitted from publications than significant results. Omission of bias adjustment has been identified as a significant problem in educational research (Ropovik, Adamkovic, & Greger, 2021). The present project is a scoping review and not a meta-analysis and as such, no statistical conclusions are drawn about effectiveness. It is exploratory in nature, attempting to encapsulate a range of perspectives about AL and explore potential constituent components

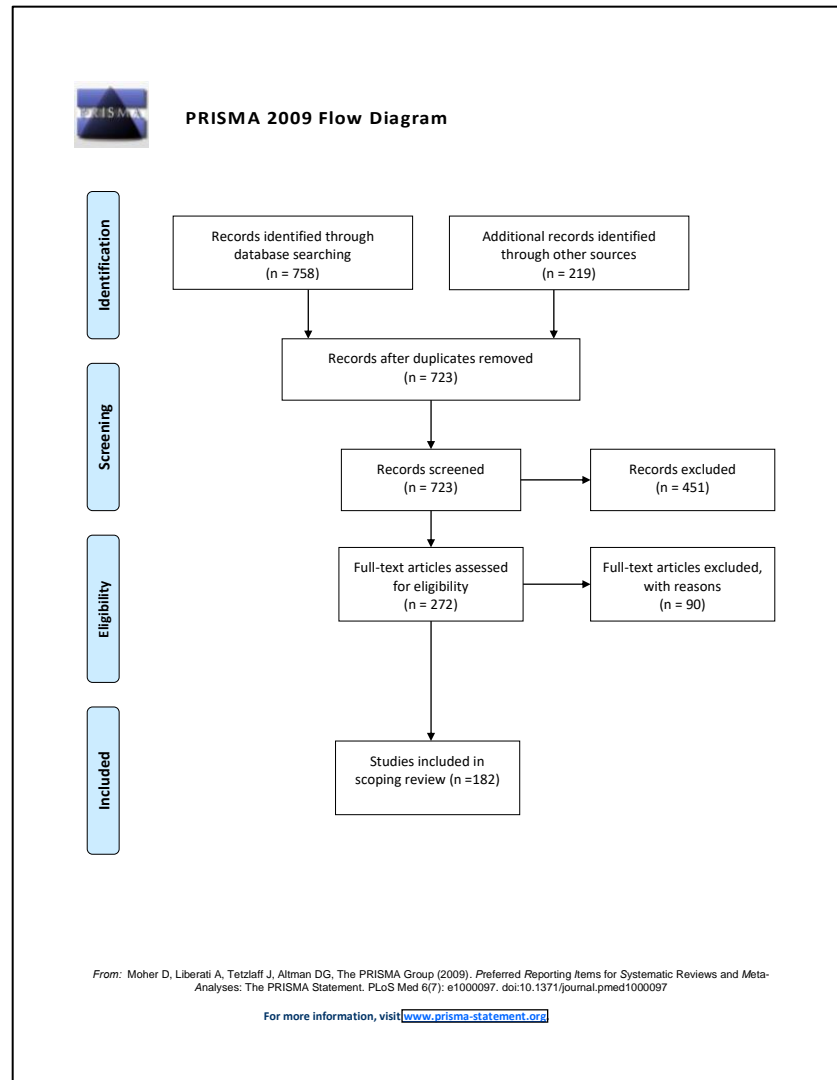
without drawing empirical conclusions. Regardless, it is important to acknowledge the potential for publication bias to influence what sources are available, what they report, and ultimately the thematic data that is extracted. It is to this end that scoping review methodology is considered appropriate; given the exploratory nature of the project, it is pertinent to begin investigating the feasibility of a unique construct of AL for HE by sweeping for research that is available in a systematic way without making claims about the full representability of the collected dataset. The target construct is as-yet unspecified, and as such any attempts to review it will likely begin an imprecise.

To identify studies for inclusion, the following search strategy was implemented for each database. Subject headings and key words for each concept were combined using the 'AND' operator. In cases where the number of results were not sufficiently reduced, an exclusion term was applied using the 'NOT' operator. For each search, results were screened by title and abstract. As with the prior phase, results had to reference higher education as the research context, as well as conceptual discussion related to assessment, to be considered for inclusion. Selected entries were extracted to reference management software and categorized according to date, database, and search term. These database searches resulted in 758 journal articles, book chapters, theses and dissertations. Next, the reference lists for each of these entries were perused for additional sources. DOI links were used to locate these papers; where DOI's were broken or not available, reference data was searched using the internet search engine Google. In the reference list search, 219 additional sources were identified. These studies were evaluated using the same criteria as in prior phases.

After consolidating entry information and removing duplicates from the library, a total of 723 sources were entered the title and abstract screening process. The information from this list was organized in open-source reference management software. Columnized data included date, authors, titles, and abstracts. At this stage, two reviewers screened the library by reading the title and abstract and rating each paper as YES, MAYBE, or NO. The reviewers were the lead researcher of the present thesis and the supervisor of their graduate program. Instances in which both reviewers rated the paper as YES resulted in extraction to the next stage. Cohen's kappa was used to calculate interrater agreement, resulting in $k = 0.8172$ which can be interpreted as substantial to almost perfect agreement (Landis & Koch, 1977). In total, 272 entries were agreed upon by the raters as fitting the criteria for inclusion in the scoping review. Next, the full text for each entry was read and assigned a rating from 1 to 4, where 1 indicated irrelevance and 2, 3, and 4 indicated low, moderate, and high relevance. This rating was intended to guide the inductive thematic coding: papers would be read and coded in order from high-to-low relevance to discern themes from papers with higher relevance and thus increase visibility of these codes in lower-relevance papers. After reading each full text, 54 papers were rated 1 and excluded from the final analysis. An additional 36 papers were removed due to inaccessibility or non-obvious duplication. For example, in some cases the author of an included dissertation had published a journal paper based on their dissertation research, which was also included. In these cases, the record was consolidated into one. In total, 182 papers were retained for thematic analysis. All

excluded sources were deemed to not fit the inclusion criteria, or else to explicitly meet the exclusion criteria as described previously. Most frequently, sources were excluded due to targeting populations other than HE faculty. Figure 3.1 visualizes the selection and reduction processes using a PRISMA flow diagram (Moher et al., 2009) to report final numbers:

Figure 3.1: PRIMA scoping review summary.



Additionally, summary information was collected from each article, including purpose, type of study, characteristics of participants, definitions of AL for HE (if any), and results.

The library of 182 included papers was imported into scoping review management software for thematic analysis. A general inductive approach (i.e., Braun & Clarke, 2006; Thomas, 2003; Mykota, & Muhajarine, 2005; Hootz, Mykota, & Fauchoux, 2016) was used to determine the themes and their respective subthemes emerging from the studies. For each paper, thematic analysis occurred as follows. In the first phase of coding, the full text of the paper was

read, and key thematic phrases were annotated. Components stated or implied as relevant to AL were charted into Microsoft Excel, and prominent themes for each paper were summarized and categorized. Each component code was applied to a paper when it discussed that component in a fashion suggesting it as important for effective HE assessment. As such, any given source could have any number of component codes. This allowed tracking of how frequently each speculative component was referenced by each paper within the scoping review library.

In the second round of coding, tabulated components were manually clustered based on subjective similarity. For example, codes such as ‘understanding’, ‘knowledge’, and ‘expertise’ were clustered in one column labeled “knowledge and understanding”, whereas components like ‘practice’, ‘implement’, or ‘use data’ were clustered in another column labelled “assessment skills and practice”. Finally, all sources were read again and re-examined for the presence of all existing component codes. Such codes and their analysis are described in the following ‘ Scoping Review Results and Analysis’ chapter, and the full table of included studies sorted by theme can be found in Appendix A.

CHAPTER FOUR: SCOPING REVIEW ANALYSIS

This chapter presents the results and analysis from the scoping review study and the subsequent survey study. The scoping review results are described in terms of thematic labeling, and discussion regarding extracted components of AL for HE is interwoven throughout. The survey study results are described with regards to data collection and cleaning, statistical analysis of quantitative data, and thematic analysis of qualitative data. Discussions of the analyses are interwoven throughout the section.

4.1 Thematic Analysis

A general inductive approach was used to explore themes and subthemes emerging from the scoping review data. The scoping review data was examined for thematic components following the analytic framework proposed by Braun and Clarke (2006). This involved a multi-stage thematic review in which the full text of each source was read multiple times. First, the full text for each source was read and examined by looking for examples of components of AL as implicated by the author's consistently recurring reference. Each potential component is listed and accounted for in an open coding process; following this, components were recoded axially by reorganization into categories based on context (Braun and Clarke, 2006). Table 4.1 demonstrates the groups of component codes that emerged from thematic analysis and the number of papers that referenced each code group:

Table 4.1: Components and Components Groups

Component Code	# papers
Knowledge & Understanding	109
Assessment Skills and Practice	97
Attitudes, Beliefs, Conceptions	72
Engagement and Experience	63
Purposes of Assessment	52
Community/Collaboration	52
Contexts of Assessment	49
Decision-making	40
Assessment Fluency and Dialogue	37
Local Context and Resources	37
Metacognitions and Reflection	35
Response to External Demands	23
Integration in Pedagogy	17
Comprehension of Standards	16
Demographic Influences	9
Leadership and advocacy	5

This phase of coding resulted in sixteen clusters of codes seeming necessary for AL for HE: Knowledge and understanding (109), skills/practice (97), conceptions (72), engagement (63), local processes (55), community (52), purposes (52), decision-making (40), dialogue (37), context (37), self-regulation (35), response to external pressure (23), pedagogy (17), standards and outcomes (16), demographic influence (9), and assessment leadership (5).

In the second read of the full texts for thematic analysis, each source was read again and assigned a thematic statement informed by the component codes, which were recontextualized to demonstrate major themes (Braun & Clarke, 2006). These statements were organized into like groups, and then each source was assigned a more comprehensive thematic statement according to group membership. Each unique theme was recorded in a list and categorized with similar themes. The extracted information was entered into MS Word files with information synthesized, tabulated, and summarized by content. After this process, the themes were recontextualized as follows:

Table 4.2: AL for HE Codes to Themes

Scoping Review Code	Scoping Review Subthemes	Thematic Category
Knowledge and Understanding, Purposes of assessment, comprehension of standards	Knowledge base for assessment	Stable or Explicit Knowledge
Skills and Practice, integration in pedagogy	Disciplinary knowledge, SBA/SLO assessment, student assessment, knowledge of stakeholders	Flexible/Dynamic Knowledge
Metacognitions and reflection, response to external demands	Understanding promotes participation, metacognition,	Self-regulating Knowledge
Attitudes, Beliefs, Conceptions	Factors shaping conceptions,, beliefs, cognitions, limiting conceptions	Conceptions of Assessment
Fluency and Dialogue	Fluency with assessment language,	Fluency with Assessment Language
Community/collaboration, Contexts of assessment, Local Resources	Contexts, cultures of assessment, institutional characteristics,	Local Contexts
Decision-making	Compromise of competing demands in context, Using data, judgement of evidence	Decision-making
-	Development, Institutional support, Development in practice, need for development	Developing Assessment Literacy
Engagement and Experience, Demographic influences, Leadership, and advocacy	Professional role, experience and expertise, ownership	Engagement/Assessor Identity

The thematic analysis resulted in eight major AL themes with some contained subthemes (full thematic table in appendix A). First was the theme of conceptualizing assessment literacy, with the subtheme's external evaluation, LAL, leadership, and classroom AL. Knowledge and practice for HE assessment contained the subthemes fluency in assessment language, knowledge base for assessment, disciplinary knowledge, understanding promotes participation, and SLO/SBA. Conceptions of Assessment contained the subthemes factors shaping conceptions, beliefs, cognitions, and conceptions limiting AL. Context contains the subthemes' professional role and cultures of assessment. Development contains the subthemes Institutional support for faculty AL development, need for development of AL in HE, and literacy develops in context in practice. Decision-making contains the subthemes compromise of competing demands in context, experience and expertise, and judgement of evidence. Self-regulation contains no subthemes. The theme various influencing factors contains the subthemes institutional characteristics, ownership, and cultures of assessment.

For the third and final read of each full text, extracted papers were sorted to isolate papers with a focus on professional assessment tasks in HE as opposed to a focus on student learning assessment in HE (Appendix B). Using this thematic organization, 42 papers were isolated discussing assessment for the following professional tasks: external evaluation (5), language development (12), academic development (3), disciplinary professionalism (12), program assessment (4), assessment culture (3), institutional planning (1), and faculty evaluation (1). This information was used to understand how HE AL is discussed in academic literature as pertaining to non-student assessment tasks.

This follow sections begin by addressing the first research question by briefly describing each extracted theme and subtheme. These descriptions are given structure by comparing extracted themes to an existing conceptualization of AL for K-12 contexts (Teacher Assessment Literacy in Practice: Xu & Brown, 2016). Next, themes from the subgroup of papers relating to specific professional focuses for AL in HE will be summarized in response to research question two. Finally, implications from the scoping review are discussed to establish research questions for a subsequent follow-up questionnaire to explore how HE faculty implicate AL in their everyday work.

4.2 Research Question One

What are the elements and features of "Assessment Literacy" in higher education that set it aside from school-based (Pre-K through 12) education?

4.2.1 Conceptualizing AL in HE

At the time of this review, no sources had conceptualized AL in terms of generalizable elements and features for effective work in HE. However, AL has been explored in relation to specific disciplinary and professional tasks including: external evaluation, where assessment communities are considered key influences on practice; assessment leadership, where competencies in assessment skills are necessary for those coordinating assessment efforts;

language assessment, where core competencies of applied assessment knowledge have been highlighted for practice and classroom assessment; and classroom assessment, considered a broad domain where assessment skills enhance teaching and student learning. Generally, Medland (2015, 2016, 2019) described six facets of AL for external evaluators in HE, highlighting the impact of local assessment contexts on individual AL. These components assumed AL as a socially negotiated and personally developed variable rather than binary (i.e., present or absent) and called for further work to understand AL in HE contexts. These assumptions are common among conceptions of AL from other disciplines (e.g., LAL: Kremmel & Harding, 2020). This highlights the need to explore what professional tasks might be relatively constant among HE faculty, how assessment is involved with these tasks, and what elements of AL may be necessary for effective assessment. Elements and features of these assessment competencies are further explored later in this chapter in relation to professional tasks in HE involving assessment.

Outside of these professional conceptualizations, it is pertinent to compare how AL has been conceptualized for school-based education to verify elements and features of AL for HE as emergent from this review. A scoping review conducted by Xu & Brown (2016) examined literature relating to assessment literacy and teacher education to examine the specific assessment literacy of teachers. Their review resulted in three broad thematic categories: knowledge and practices, with subthemes knowledge base, measuring teacher AL, and AL measurement validation; Assessment education and its relationship with various mediating factors, with subthemes assessment courses, assessment training programs and resources, relationship among assessment training, teacher conceptions of assessment and AL, and teacher assessment training needs and self-reported efficacy; and contextual consideration of AL, with subthemes macro and micro contexts, teacher's identity as assessor, and understanding and developing AL in practice. Data from the scoping review informed the development of a model for Teacher Assessment Literacy in Practice (TALiP), which attempted a comprehensive approach to assessment literacy applicable to teachers in school-based education.

For the present analysis, scoping review themes were compared to the TALiP model to examine how AL elements and features unique to HE settings may be made salient. This comparison helped to organize scoping review themes into a framework to support future survey development. For this comparison, extracted thematic categories were loosely reorganized to match Xu & Brown's conceptual structure as in Table 4.2:

Table 4.2: Thematic comparison to TALiP.

Identified Themes of AL for HE	Xu & Brown's Components: TALiP K-12
Knowledge and Practice for HE Assessment <i>Knowledge base for assessment</i> <i>Disciplinary Knowledge</i> <i>SLO/SBA</i> <i>Fluency in Assessment Language</i>	Knowledge Base <i>Disciplinary knowledge and Pedagogical Content Knowledge</i> <i>Knowledge of assessment purpose, content, and methods</i> <i>knowledge of grading</i> <i>Knowledge of feedback</i> <i>Knowledge of assessment purposes, content, and communication</i> <i>Knowledge of Student involvement in assessment</i> <i>Knowledge of assessment ethics</i>
Self-regulation <i>Understanding promotes participation</i>	interpretive and guiding framework <i>beliefs affecting how knowledge is 'taken up'</i> <i>knowledge that is deemed to be useful</i>
Conceptions of Assessment <i>Factors shaping conceptions</i> <i>Beliefs</i> <i>Cognitions</i> <i>Conceptions limiting AL</i>	Teachers Conceptions of Assessment <i>Affective dimensions</i> <i>cognitive dimensions</i> <i>views of learning and epistemological beliefs</i>
Context <i>Cultures of assessment</i> <i>Institutional support for Faculty AL</i> <i>Institutional Characteristics</i> <i>Ownership</i>	Micro- and Macro- Contexts as Boundaries <i>Macro- sociocultural</i> <i>micro-institutional</i>
Decision-making <i>Compromise of competing demands in context</i> <i>Experience and Expertise</i> <i>Judgement of Evidence</i>	Teacher Assessment Literacy in Practice <i>Compromises among tensions</i> <i>Compromises in decision-making</i> <i>compromises in action taking</i>
Development <i>Institutional support for faculty AL</i> <i>Need for development of AL in HE</i> <i>Literacy Develops in Context in Practice</i>	Teacher Learning <i>reflective practice</i> <i>participation in community learning activities</i>
HE Role Enactment (?) <i>Professional Role</i> Note: This table is colour-coded to identify comparable themes from each thematic approach.	Teacher as Assessor <i>reconstruct identity to include role as assessor</i>

4.2.2 Knowledge and Practice for HE Assessment.

Xu and Brown's (2016) component of knowledge shows practical elements tending towards specificity of assessing students, which reflects the professional foci of school-based teachers. In comparison, the subtheme knowledge base also emphasizes appropriate awareness, understanding, and implementation of empirically supported techniques (Adachi, Tai, & Dawson, 2018; Beebe, Vonderwell, & Boboc, 2010). Papers demonstrating the thematic component of knowledge and practice for HE assessment exemplify a wide range of assessment-related theory, skills, and competencies deemed important for effective assessment efforts in HE. A knowledge base for assessment is considered essential to underpin assessment practice, such as an appropriate understanding, awareness, and implementation of empirically supported techniques (Adachi et al., 2018). Understanding of assessment theory is necessary to guide appropriate assessment design as practice without sufficient knowledge reinforces cycles of inefficiency (Dietrich, 2011; Melguzzio et al., 2014). Such assessment understanding has elements of underlying stability consistent across contexts (Beebe et al., 2010), but also awareness that best assessment practices reflect specific purposes in particular contexts (Norcini et al., 2011).

Another subtheme is variation in assessment knowledge and practices among disciplines. Sources reflecting this subtheme acknowledged the significant subjective difference in how assessment is enacted among academic disciplines (Fletcher et al., 2011; Jeong, 2013; Malone, 2013; Rawlusk, 2016). They take a constructivist stance of constructing assessment knowledge as necessary for particular purposes in particular disciplines. Fletcher et al. (2011) reason that knowledge bases constructed as bound to specific disciplinary content are important to demystify assessment by making processes tangible and discussable. Of course, this subtheme is relevant to the initial problem of this thesis: that HE assessment and AL are not well conceptualized, resulting in variation and misunderstanding among uses and contexts. These subthemes thus reveal a tension between the practical value of a standardized conceptualization of AL versus a contextually variant understanding of assessment from a subject-oriented frame. For example, sources reflecting this subtheme recognize a relationship between disciplinary assessment knowledge and other aspects of AL in HE, including assessment task selection (Goubeaud, 2010; Harland et al., 2015; Swarat et al., 2017), faculty engagement with assessment (Cole & De Maio, 2009; Hines, 2009; Hutchings, 2011; Pawlyshyn, 2013), or conceptions of assessment (Jeong, 2013; Malone, 2013; Swarat et al., 2017).

Tensions regarding the usefulness of top-down versus ground-up approaches to assessment understanding are also explored in this subtheme. Poor individual understanding of assessment is continually referenced in relation to resistance or disengagement (Cole & De Maio, 2009; Hines, 2009; Macdonald et al., 2014; Marrs, 2009; Tovar-Klinger, 2016). Cole and De Maio explain that when faculty are unfamiliar with assessment, they may view it more negatively; thus, sufficient and specific understanding of assessment theory may mediate reduced resistance to assessment processes in HE. Similarly, Cohen (2004) found that knowledge of assessment theory improves conceptions, assessment implementation, and thus decision-making,

arguing that faculty development is necessary to support a sufficient breadth and depth of theoretical understanding. Bandy et al. (2016) also found that ‘good’ assessment occurs when involved faculty have sufficient understanding higher-order assessment purposes, whereas Macdonald et al. (2014) cautioned that resistance may occur when assessment purposes are not understood. Summarizing these sentiments, Pawlysyn (2013) explained that AL informs engagement, which in turn enables new assessment approaches and informs assessment at different levels. Clearly, any conceptualization of AL for HE must include a theoretical base of *a priori* theoretical knowledge while also promoting sufficient contextual understanding of assessment as bound to particular disciplines.

4.2.3 Assessment Fluency

In these scoping review findings, significant attention was paid to meticulous clarity in usage of assessment terminology, which is justified given the broad variety of contextual factors involved in HE assessment such as disciplinary traditions or institutional regulations (Richards & Pilcher, 2014; Taras & Davies, 2014). With no direct comparison in Xu & Brown’s (2016) model, the importance of mutual understanding of assessment language emerged with relative prominence in this review compared to TALiP. It involves moving beyond understanding assessment personally into establishing a mutual understanding of assessment in any given assessment interaction (Forsyth et al, 2015; Medland, 2016). It could be inferred that such diligence to language, though important, is less critical in school-based education as there is less diversity among stakeholders and professional purposes of assessment. In higher education contexts, even student assessment efforts are subject to departmental, disciplinary, and overall contextual variation. As well, there exists a wider range of stakeholders in HE assessment, whose personal assessment understandings are bound to their contexts. These stakeholders with diverse understandings and various levels of literacy as applied to their own duties necessitates precise attention to language when discussing assessment so that miscommunications can be limited.

4.2.4 Metacognition

Papers with the theme of metacognition denoted aspects of how HE faculty make the impacts of underlying assessment components salient. Metacognition may be underrepresented as a thematic category given its tendency to arise in relation to other themes and subthemes, like assessment knowledge, assessment cognitions, or assessment fluency. For example, metacognitive awareness guiding one uses assessment language themselves, and to establish shared understanding in diverse settings and interactions, has been espoused within HE contexts as mark of sophisticated AL (Medland, 2018).

In the TALiP model, an interpretive guiding framework mediates how theoretical knowledge is deemed to be useful and taken up (Xu & Brown, 2016, p. 156). The component of ‘metacognition’ somewhat mirrors this metacognitive awareness in the sense that faculty demonstrating more sophisticated AL seem to engage in reflection about what they know and the purposes for which assessment is invoked (Bandy et al., 2016). As well, the subtheme of

understanding promotes participation reflects how knowledge of assessment theory may facilitate improved conceptions, implementation, and thus decision-making (Cohen, 2004). This could be compared to a feedback loop with multiple opportunities for reflection about components of AL to bolster understanding and develop AL sophistication. Here, the purpose of metacognitive feedback is reflecting upon and critiquing what is known, believed, and understood about assessment. This feedback loop may be where resistance to assessment, as the implied antithesis to ‘deeming assessment useful’, may come in to play. Resistance to assessment occurs when a limited knowledge of assessment results in assessment processes being deemed not useful (Macdonald et al., 2014; Marrs, 2009). Such resistance, paired with limited assessment knowledge, seems to create ‘rote’ HE faculty engagement with assessment processes.

4.2.5 Conceptions of Assessment

The TALiP component ‘Teachers’ Conceptions of Assessment’ contains views of learning and epistemological beliefs as well as affective and cognitive dimensions (Xu & Brown, 2016). The AL for HE theme conceptions of assessment speak to these features at a more general level, describing internal/personal dimensions shaping views of assessment such as personal variables (Alsobrook, 2010; Myers & Myers, 2015), institutional variables (Feuerstein, 2015), discipline (Dueben, 2015; Halinen et al., 2014; Hidri, 2016), or experience with assessment (DiLoretto, 2013; Ebersole, 2009). Here it remains to discern what assessment tasks are being conceived of, given how a broad portion of the literature could be framed as ‘conceptions of student learning assessment’. There seems to be a limited pool of knowledge regarding how assessment tasks *not specific to* student assessment are conceived of by higher educators.

4.2.6 Context

The TALiP model next explores micro- and macro- contexts as boundaries to practice, given that teachers as assessors are constrained by local community and broader sociocultural norms, policies, and contexts of practice. Likewise, the assessment practices and roles of HE faculty are constrained by needs and purposes and at all contextual levels. The AL for HE scoping review found themes comparable to these categories including general context, local community contexts, institutional context, and trust.

Together, these themes underscore the contextually situated nature of AL. Regardless of assessment knowledge, conceptions, and language fluency, assessment practice is constrained by contextual factors like disciplinary norms (Bloxham et al., 2016; Heinrich, 2015), local assessment communities and culture (Emil, 2011; Fuller et al., 2016; Skidmore, Hsu, & Fuller, 2018) access to resources for assessment (Creason, 2016), institutional norms and pressures, and larger societal pressures (McCune, 2018). The complexity with which these and other factors can be arranged reveals the nuance inherent in managing pressures and relationships among stakeholders of differing interests, departments, and individual actors in higher education settings. Thus, this theme of AL for HE involves an understanding of how contextual factors

create boundaries and possibilities for practice and responding to such boundaries and possibilities by enacting appropriate knowledge in communicable ways. Contextual navigation refers to a diplomatic ability of selecting and enacting the best knowledge for a given assessment task as possible within the contextual boundaries set a range of levels from micro- to macro-.

4.2.7 Decision-making

Xu and Brown's next component is titled TALiP, and rests upon the knowledge base, conceptions of assessment, and boundaries created by subjective contextual factors. It is an inherent dimension of compromise in decision-making among competing tensions whereby "teachers balance the demands of external factors and constraints with their own beliefs and values as informed by their theoretical and professional knowledge" (p. 157). Essentially, TALiP denotes the balancing point where tensions involving context, knowledge, and conceptions are balanced to meet the specific needs and purposes of various stakeholders. Xu and Brown's approach for teacher AL involves decision-making regarding student learning assessment, but for AL for HE settings included the balancing of professional purpose. While decision-making regarding student learning assessment in HE may appear like its counterpart in school-based settings, these broader assessment roles in higher education may entail high-stakes decision-making that may not be recoverable as opposed to classroom assessment. Examples of such purposes include assessment leadership (Janke et al., 2016), external evaluation (Medland, 2019), and program assessment (Emil & Cress, 2014). These and other non-student assessment tasks require the same or greater compromise than student learning assessment tasks.

4.2.8 Development

Professional teacher learning is framed as an impetus for TALiP due to the dynamic nature of classroom assessment, and how teacher questions and concerns from their classroom practice may lead to reflection and changes. AL development for student learning assessment functions in school-based education is thus motivated by participation in the dynamic assessment environment and reflection on practice. The development theme of AL for HE mirrored this with an added emphasis on how ownership, engagement, and experience seem to facilitate further development of faculty AL in HE settings (Caudle, 2014; Clark & Filinson, 2011; Deeley & Bovill, 2017). Opportunities for development have been found to improve sophistication of faculty assessment knowledge (Turner, 2013), beliefs (Kramer, 2008; Peterson, 2019), and participation (Grunwald & Peterson, 2003). Contextual boundaries for assessment are also acknowledged in research literature finding AL to develop in context and practice (Leary, 2017; Presley, 2015). In addition, there is a significant amount of literature dedicated to meso-level institutional resourcing for AL, arguing that development should target knowledge, conceptions, and ownership (Frey & Overfield, 2002; Guetterman & Mitchell, 2016; Haviland et al., 2011). Taken together, themes from the scoping review echo Xu & Brown's conclusions in that development may occur if faculty are aware of "the reciprocal interactions among many processes, mechanisms, and actions arising from assessment activities" (2016; p. 158). It seems

that AL in HE develops with assessment experiences, and assessment community membership acts to link past experiences with future possibilities (Jawitz, 2008; 2009).

4.2.9 Professional Role Enactment

The top of Xu & Brown's (2016) model is reconstruction of identity as assessor, which relocates focus from particular assessment tasks onto the individual's professional participation in assessment activity, utilizing and reinforcing all prior elements. Some papers from this review address assessor identity considering self-regulation and understanding purposes of assessment (Bandy et al. 2016). Further, the AL for HE scoping review found themes and components that did not fit neatly alongside the TALiP model including leadership, assessment advocacy, pedagogy, and scholarship. These themes illustrated various professional foci of assessment which may be approached by an individual who has constructed their role as an assessor. As such, the TALiP component of teacher-as-assessor may reflect Xu and Brown's (2016) application of AL knowledge to the specific professional domain of student assessment. The present review yielded a thematic category labelled as professional roles which did not nest within other thematic categories. To be specific, AL in HE clearly demonstrates an element of selective professional role knowledge wherein assessment knowledge, skills, are conveyed of as useful and utilized for particular but necessary tasks for some (but not all) faculty in HE contexts such as student assessment, language assessment, or external evaluation. A shift of framing is thus observed from an overarching AL into more specific professional literacies as applied to specific assessment purposes, projects, and tasks. This begs the question: For what purposes, projects, and tasks do faculty apply assessment literacy in HE contexts?

In summary, the present scoping review of AL for HE settings revealed comparable elements to Xu and Brown's TALiP model, and as such it is possible that AL for HE faculty develops along a similar continuum of sophistication. As expected, however, sources in the dataset were largely focused on student learning assessment in HE. The importance of student learning assessment cannot be downplayed, and thus it has been important to examine scoping review data with consideration for how student learning assessment remains a significant professional responsibility for HE faculty. In order to achieve the aims of this thesis, assessment in other forms ought not to be diminished by historical emphasis on student learning assessment. It is pertinent, then, to turn attention in HE to the different professional roles wherein AL is necessary.

4.3 Research Question Two

A second research question asked how these elements could be integrated in a conceptual model to develop a useful unified definition of AL for HE contexts. The scoping review investigated AL attempting to determine elements and features specific to higher education settings and contexts. It specifically identified a significant body of literature related to knowledge and competencies needed for assessment processes supporting student learning, such as: classroom assessment, SLO/SBA, formative/summative assessment, student assessment

design, conceptions of student assessment, or decision-making using student assessment data. Research on these assessment competencies evoked AL as it has been conceptualized at micro-levels for school-based education, which was analyzed through comparison to Xu & Brown's (2016) model of TALiP. However, these themes continue to neglect features of AL pertaining to professional duties and tasks in higher education that are not present in school-based education. Key to this idea, Arreola, Theall, and Aleamoni (2003) succinctly described professional activities like conducting research, delivering faculty development programs, and designing faculty evaluation systems where HE faculty "must perform at a professional level in a variety of roles that require expertise and skills in areas that often extend beyond the faculty member's specific area of scholarly expertise" (p. 1). This contention holds true for HE assessment, and to such an end this scoping review determined a gap in literature related to AL necessary for professional responsibilities of higher educators not relating directly to student learning or classroom assessment. Some papers were found related to professional tasks necessitating AL, such as assessment of scholarly research, external evaluation/accountability efforts, and some processes relating to some high-stakes decision-making that may not be recoverable in comparison to student assessment. The following sections summarize themes that emerged from HE AL literature unrelated to student assessment processes. Where possible, language conventions for discussing AL for HE are used as posited by Medland (2015, 2019) as well as definitions for AL-related terms as clarified by Fulmer, Lee, & Tan (2015).

4.3.1 Quality Assurance Processes and External Evaluation

A small subgroup of papers examined assessment for quality assurance purposes and how AL is necessary for such processes. From a general and internal frame, Rosa, Sarrico, and Amaral (2012) found varying levels of faculty perceptions of support for various intended purposes of HE quality assurance including communication, motivation, control, improvement, and innovation. Academic respondents most highly favored assessment conducted for the improvement purpose, which can be translated in the present context as for "the development of their own skills and competencies, or a better link between teaching and research" (p. 360). As well, they favored assessment for communicative purposes such as developing mechanisms and processes to increase the transparency of the quality of HE systems. Less supported are perceptions of motivation and control purposes, to which resistance may be expected given the potential for such processes to impose constraints on the agency of HE faculty. Overall, these academics' knowledge and understanding of quality assessment purposes seem influenced by the degree to which they benefit and align with the local values of academics. The theme of self-regulation may be invoked as a component of AL for HE wherein awareness of ones' own perceptions and values interact with an assessment knowledge base. In this case, self-regulation describes how AL may be bolstered by supporting increased individual faculty awareness regarding their professional values and how those values may impact how they engage with assessment. This seems to be supported by Seema, Udam, and Mattisen (2016), who approached external evaluation in the context of self-determination theory and found academic staff who

perceived themselves as competent and intrinsically motivated were more likely to perceive external evaluation positively. Thus, faculty who feel confident utilizing their assessment understanding are more likely to perceive assessment for non-student learning purposes more positively, suggesting that more sophisticated levels of AL involve a knowledge base, confident implementation of that knowledge, and positive perception of various purposes for which that knowledge had been implemented.

One specific HE quality assurance process is external evaluation wherein an impartial external peer reviewer evaluates the assessment processes and adherence to standards of a given institution. Medland (2015) noted how external examiners are generally appointed to the role based on ‘subject expertise’ rather than ‘assessment literacy’, highlighting a professional role wherein advanced disciplinary content knowledge rarely precludes sufficient assessment expertise. As described previously, Medland (2018) conceptualized six constituent components of AL through a small-scale thematic analysis of the written reports of external examiners. Medland (2019) then explored these components in a small-scale interview study, elaborating on their potential utility for achieving shared discourse surrounding AL.

4.3.2 AL for Academic Developers

A small group of papers considered AL necessary for academic developers to support other HE faculty in developing AL. Hughes (2009) presented frameworks for organizing the types of assessment development initiatives undertaken by universities and the functions of academic development units, arguing that AL is institutionally supported when roles and practices are clear. Beckwitt, Silverstone, and Bean (2010) explained further how institutional resources can be used to develop faculty AL, noting that faculty ownership of and engagement of assessment processes both requires and facilitates AL. This idea reconnects with the theme of self-regulation espoused by Medland (2018) and was exemplified in the present scoping review by adhering to the ‘feedback loop’ of reflection, engagement, and fluent assessment language use bolstering AL. To this end, Reder and Crimmins (2018) assert how active partnership between HE faculty and academic developers enables appropriate use of assessment data for faculty development and empowers faculty ownership.

4.3.3 Language Assessment Literacy

AL for language testing, abbreviated in that field as LAL, has been developed with sophistication by language researchers; efforts from that domain may be generalizable to AL for HE settings. This review collected some of this work as it pertains to HE contexts, though some influential papers were excluded due to nonrelevance to HE settings. Early development involved advocacy for a view of AL specific for language assessment purposes, such as Inbar-Lourie (2008) who proposed components of a LAL knowledge base. O’Loughlin (2011, 2013) echoed this call, elaborating that LAL is necessary for language test users in HE (such as researchers or admissions officers) to use language tests appropriately, interpret test data correctly, and make just decisions using language data. The impact of discipline and personal

factors on how LAL knowledge is utilized was posited by Jeong (2013); meanwhile, Malone (2013) drew distinctions between the beliefs of different stakeholders in language assessment processes, noting that language testing experts consider tests differently than ‘test users’ such as language instructors. This emphasis on different stakeholders was emphasized by definition of LAL as a profile of competencies, variable to the needs of an individual’s professional context, rather than a knowledge base (Taylor, 2013; Baker, Tsushima, and Wang, 2014; Baker, 2016).

Similarly, Kvasova and Kavytska (2014) expanded responsibility for language professionals to develop not only their knowledge base, but also understanding of purposes, assessment design, and implications of language assessments. A scale was developed by Ölmezer-Öztürk and Aydın (2018) for measuring language instructor’s LAL, and emphasis was re-drawn to the need for LAL development among HE language professionals by Dibiaselubrano (2018). Most recently, particular components of AL are being elaborated upon. Deygers and Malone (2019) examined conceptions affecting how LAL knowledge is understood and interpreted in specific professional contexts, finding that beliefs and practical considerations seem to override empiricism, policy considerations, and even awareness of best practices. Kremmel and Harding (2020) approached empirical conceptualization of the profile model of LAL using factor analysis, extracting nine components which load on to a higher-order factor of LAL: Developing and administering language assessments, assessment in language pedagogy, assessment policy and local practices, personal beliefs and attitudes, statistical and research methods, assessment principles and interpretation, language structure, use and development, washback and preparation, and scoring and rating. Taking these sources together, LAL has seen considerable attention from language researchers

Given the relative efficacy with which LAL has developed, what can be learned about conceptualizing AL for HE from the language testing disciplinary community? First, it seems that the call for conceptual work seems to have been approached by the community with relative rigor and structure. Taylor’s (2013) seminal conference paper structured LAL conceptualization through a series of research areas devised by the language research community at a national conference, including investigation of key stakeholder groups, necessary content input, specific domains and contexts, and timeliness and structure of research. An internationally and methodologically diverse group of papers were produced in response to these areas of investigation and have established the groundwork for what has developed into contemporary, thorough, and useful conceptualizations of LAL. In comparison, the present review underscores Medland’s (2019) assertion that papers addressing HE contexts have tended to delineate the term AL differently with respect to language, rules, standards, knowledge, skills, and attributes. Indeed, DeLuca, LaPointe-McEwan, and Luhanga (2016) identified that the meaning of AL in school-based contexts shifts to meet the needs of the research focus; the research identified in the scoping review affirms this for HE settings.

Perhaps the difficulty in conceptualizing AL for HE has arisen due to the broad swathes of HE stakeholders using assessment for diverse and particular functions, and thus the high impossibility of establishing a coordinated conceptualization effort crossing contextual

boundaries such as discipline. The domain of language testing can be clearly defined, with clear roles and functions for specific stakeholders, whereas the boundaries of higher education are broad, multifaceted, and related in complex ways – thus, struggles to define AL for HE may arise from this complexity. The domain of educational assessment research serves as a prime example. As noted by Taras and Davies (2018), functions of student learning assessment have been falsely and problematically dichotomized into the bipolar categories of more-desirable formative assessment and less-desirable summative assessment. As well, Medland (2019) summarized how such research tends towards deficit models in their description of AL as present (or high, or sophisticated) or, more likely, as absent (or low, or unsophisticated). Given how these popular conventions force assessment to be categorized according to student learning functions, it is no surprise that AL research in HE has proceeded using the uncontested assumption of assessment as primarily *of*, *for*, and *as* student learning. These studies are useful in drawing attention to the need for AL development in HE settings but are less useful in providing a scaffold for such development to occur. For example, some studies find low AL among their faculty, implement a seminar, course, or workshop to enhance it, and report higher levels of AL following such a workshop (e.g., Deneen & Boud, 2014; Forsythe et al., 2015; Haviland et al., 2010).

However, these studies tend to be small and bound to contextual variables such as department, institution, or nation. Indeed, some have found that AL improvement from such workshops does not persist and may in fact worsen over time (Haviland et al., 2011). From the example drawn from LAL, it would seem that attention is better directed towards discovering the AL needs of stakeholders in HE and developing a conceptual model that is flexible to those needs. The profile of competencies approach recently refined by Kremmel and Harding (2020) seems to have been useful in ensuring LAL is applicable to community members different contextual settings conducting and using language tests for different purposes. Practical utility seems to override adoption of more strict empirical boundaries, and as such adopting a model of LAL that accounts for the needs of various stakeholders may contribute to the conceptual success of LAL thus far (Deygers & Malone, 2019). A model of AL for HE, then, must address the specific needs of stakeholders who use assessments while performing various roles and functions within HE settings. This convention may again explain why AL research for HE has gravitated towards student learning assessment as a relatively ubiquitous responsibility for a large swatch of academic faculty. However, it also provides an avenue to ensure future AL research may simultaneously address the broad complexity of assessment in HE and the context-situated nature of individual assessment practice.

Finally, the empirical components extracted by Kremmel and Harding's (2020) recent factor analysis of LAL can be examined in comparison to the broad data from the present review. First, it is apparent that only one of their nine factors (Language structure, use, and development) is purely domain-specific content knowledge – it is concerned with aspects related to language and not assessment-related competencies. Two more factors (Developing and administering language assessments, and Assessment in language pedagogy) could be considered discipline-specific forms of factors that could be generalized by dropping the 'language' requirement (i.e.,

transforming to ‘developing and administering assessments’, and ‘assessment in pedagogy’). Five of the remaining six factors are more generally concerned with aspects of assessment and align with themes and components emergent from the present review. Taken broadly, Kremmel and Harding’s (2020) extracted factors may be exemplar of how a more general form of AL for HE is applied to the specific professional domain of language testing. It is worth noting how three of the components (Assessment in language pedagogy, washback and preparation, and to a degree, scoring and rating) reflect the professional role of student instruction, whereas others (e.g., statistical and research methods) may be utilized less by such faculty whose primary role is language instruction. Thus, Kremmel and Harding (2020) achieved a model representing general features of HE assessment as well as profession-specific roles and responsibilities, while remaining useful by accounting for various levels of each competency needed by different stakeholder groups. For the aims of the present study, it may be useful to investigate to what degree these factors generalize to a broader conceptualization of AL in HE settings.

4.3.4 Professional Roles and Assessment Literacy

Much of the remaining literature from this review that did not deal directly with student learning AL instead addresses several aspects of assessment professionalism for HE. Some of this literature has been referenced already, such as Arreola, Theall, and Aleamoni (2003) who frame professorial responsibilities such as AL as *meta-professional* in the sense that they lie outside a given faculty member’s area of content expertise but remain necessary for the effective performance in HE settings. Similarly, Holroyd (2000) had called for assessment as playing a key role in the construction of academic professional role, and particularly as the ‘academic-as-educator’ role gains prominence. Though relating essentially to student learning assessment, Holroyd asserted the necessity of a wider range of knowledge and a supportive institutional environment to improve assessment practice among HE professionals. More recently, Norton, Floyd, and Norton (2019; Norton et al., 2010) explicitly connected absence of faculty AL to a lack of opportunities to understand assessment practice as professional work. They investigated academics’ assessment design practice and found evidence for professionalism as well as practical constraints to enacting best practices, akin to Deygers and Malone (2018).

Other facets of AL for professional roles summarized in this review included the necessity for simultaneous ‘top-down’ and ‘ground up’ approaches to AL development (Bandy et al., 2016; Feuerstein, 2015), the impact of discipline-specific professional influences (Blumenstein, 2015; Ewell et al., 2011; Ion & Cano, 2011), and the need for congruence of personal faculty assessment conceptions with the local assessment culture (Emil & Cress, 2014; Emil, 2011).

4.4 Scoping Review Summary and Theoretical Framework

To situate AL within the professional context of HE, it is imperative to first account for how assessment is differently construed in HE. This necessitates several points of understanding. First, it seems necessary to understand the purposes for which AL is necessary in HE, including what assessment tasks are involved for those purposes and how these tasks are approached by

HE professionals. It further important to query how these tasks compose professional roles or duties that are assumed by HE professionals. Following this, a contextual AL for HE may account for what knowledge, skills, and conceptions are necessary to approach these tasks with efficacy. Only once these points of understanding are achieved may it be appropriate to investigate what of this professional AL is actually possessed by HE faculty and professionals. The present scoping review identified that this latter point is what has been targeted by the majority of AL in HE literature. This is problematic because many of these basal points of understanding seem to have been overlooked in research (apart from the LAL research body, who approached the topic with a similar set of questions (Taylor, 2013)). As such, much of the existing research literature on AL for HE centers on student learning assessment purposes and tasks, neglecting assessment involvement in the everyday work of HE faculty.

As a broad summary, the scoping review identified gaps in literature related to the non-student assessment tasks performed by HE faculty. The concept of AL as applied to HE settings has thus far tended to focus on the competencies needed in processes involved in assessing students, such as: classroom assessment, SLO/SBA, formative/summative assessment, student assessment design, conceptions of student assessment, or decision-making using student assessment data. Going forward, it is necessary to establish an understanding how assessment is involved in the professional tasks of HE faculty. To accomplish this, a second phase of the study investigated the following research questions with an exploratory mixed-methods survey of HE faculty in Western Canada:

- 1) What tasks are HE faculty performing in their everyday professional contexts that involve assessment?,
- 2) How can common HE assessment tasks be organized and labelled?,
- 3) What relationships exist between common HE professional tasks and theorized components of AL for HE? And,
- 4) How is HE assessment and its involvement in HE professional work uniquely construed by HE faculty?

These research questions were explored through the development of a survey guided by the thematic framework established by the scoping review study. The survey methodology is described in Chapter Five, and the results are described in Chapter Six.

.

CHAPTER FIVE: SURVEY METHODOLOGY

This chapter describes the methodology involved in the survey study. It begins with a brief justification for the survey itself followed by description of the survey development process, including content generation and expert review. Next, a description of survey participants is provided including recruitment strategies. The survey deployment procedure is then described as well as the plan for data analysis.

5. Survey Development

There were two aims to develop a survey for faculty to identify professional tasks which involve assessment. First, the survey would need to comprehensively reflect a range of popular approaches to assessment perspectives while resisting the tendency of the discipline to prioritize student assessment. Second, the survey items would need to broadly reflect diverse assessment needs of interdisciplinary HE faculty. This resulted in a survey development phase lasting about a year. A broad approach to survey development was adopted due to the wide range of beliefs, perspectives, and opinions on AL in HE. The survey was developed following guidelines for scale development as described by DeVellis (2017). As an initial theoretical model, it was determined to follow a process akin other specific conceptualizations of AL. Specifically, Kremmel and Harding's (2020) conceptualization of (LAL) accounted for various professional roles assumed by higher educators wherein LAL may be necessary but in varying arrangements. This approach is relevant to the present thesis which investigates components of AL for HE with a pragmatic focus on the everyday activities of HE faculty. Usefully, LAL is a specification of AL relevant to assessment in HE; this aligned with the goal of distinguishing a conceptualization of AL that accounts for functions of assessment other than student learning assessment. Furthermore, adhering to a dimensionality reduction procedure akin to Kremmel and Harding (2020) would allow for comparison of components from the scoping review to the opinions of higher educators about tasks which frequently involve assessment in HE.

A guiding theoretical framework for scale development was developed in the prior scoping review. DeVellis (2017) argues that a theoretical statement is imperative to inform scale development. To this end, the following research informed definition statement was compiled to guide scale development:

“Assessment Literacy is a collection of related and contextually situated competencies necessary for implementing assessment-related initiatives (DeLuca et al., 2018; Fulmer, Lee, & Tan, 2015; Willis, Adie, & Klenowski, 2013). In higher education, Davies and Taras (2018) describe how “*assessment literacy*, in any given setting, may be defined as an understanding of the issues, general and specific criteria, and standards which may enable a given individual to communicate efficiently with individuals in a similar context and also to negotiate meaning (coherently) from an informed position, on assessments of processes or products made within that context” (p. 475-6).”

This definition accounts for much of the theoretical development that emerged following the scoping review and situates the target construct of AL as a micro-contextual facet of HE assessment practice as per Fulmer, Lee and Tan (2013). Further, it situates the effort to understand AL for HE from the frame of HE faculty as acting subjects who use assessment in their everyday work as per Jurczyk, Vob, and Weihrich (2016).

As with the scoping review, this survey ought to be clearly emphasized as an exploratory effort to gather information about how faculty perceive and use assessment in relation to thematic components of AL. The final survey ought not to be considered a complete or empirical instrument, and usage of it for replication purposes would be inappropriate. However, the final survey has merit as an exploratory pilot into the feasibility of further exploration about how faculty use assessment in HE and what they need to know to become effective assessors in various roles.

5.1 Development of the Item Pool

The first phase of survey development was to write items reflecting components and themes of AL in HE as they emerged in the scoping review. The finalized survey responded to feedback from reviewers, and as such much of the discussion below does not represent the finalized survey. Despite the evolution of the survey based on review and feedback, this ‘version 0’ is discussed to illuminate how the final survey was initialized.

A first survey section gathered data on respondent demographic information including age group, sex, years working at HE level, rank at institution, and areas of research.

In a chapter on guidelines for scale development DeVellis (2017) recommends an item pool three to four times the size of the final scale. For this probing survey a scale of ~50 items was targeted as a longer survey was expected to fatigue potential respondents. As this survey was an exploration of tentative empirical support for the thematic components extracted from the review, it was desirable to have a short, efficient survey that could easily probe the perspectives of HE faculty. Second, the scale was targeting three to five useful items for each of the eight major themes elicited from the scoping review, items representing opinions about tasks for which assessment is used in HE. For this number of items it was expected that ~50 items would represent each of the themes elicited in the review.

To begin structuring the survey, it was intended to query what knowledge faculty believed to be necessary aspects of assessment in HE. This section adhered to the theoretical framework to conceptualize AL for HE as developmental, multidimensional, and profiled to different stakeholder groups. This section had two stems: “How knowledgeable do people in your faculty/department/discipline need to be about each aspect of assessment below?”; and, “How much is assessment involved with each professional role or task below?” The purpose for having two stems was to elicit faculty opinion about what AL was needed compared to what tasks were involved in AL. The response scale was Likert-style ranging from 1-5 with labels “Not at all”, “Slightly”, “Moderately”, “Very”, and “Extremely”. A Likert-style scale was

chosen to quantify opinion about how much a faculty member needed to know about assessment without forcing respondents to select extremes.

Following a process effective from a similar context, components from Kremmel and Harding's (2020) LAL conceptualization were first adapted to the to the proposed categories emergent by incorporating them alongside similar components from the scoping review. Similar items were also brainstormed to account for components of HE for AL that did not align with components of LAL. Table 5.1 shows how components specific to the domain of language assessment were initially generalized and brainstormed:

Table 5.1: Scoping Review Code Item Pool

LAL component	Reframed component	Number of items adapted
Developing and administering language assessments	Developing and administering assessments	14
Assessment in language pedagogy	Assessment in pedagogy	6
Assessment policy and local practices	Assessment policy and local practices	11
Personal beliefs and attitudes	Personal beliefs and attitudes	6
Statistical and research methods	Statistical and research methods	4
Assessment principles and interpretation	Assessment principles and interpretation	4
Language structure, use and development	-	-
Scoring and rating	Scoring and rating	3
-	Washback and preparation	4
-	Self-regulating	4
-	Developing assessment literacy	6
-	Assessment professionalism	4

Items written in each category reflected potential aspects of HE work for which assessment may be necessary; for example, items in *Developing and Administering Assessments* subscale included items like “Accommodating assesses with disabilities”, “designing scoring and rating scales”, and “writing good quality items/tasks for assessments”. Some subscales had much more items adapted due to the number of existing items from Kremmel and Harding's (2020) scale. At this phase of writing, a minimum of three to five items was targeted for each subscale. Items were generated by consulting existing AL survey instrument literature and brainstorming per guidelines posited by DeVellis (2017).

After adapting items from Kremmel and Harding's (2020) LAL scale, further items were written to specifically query how faculty respond to statements about assessment across hypothesized AL domains. The categories were developed by synthesizing scoping review component codes and subthemes to arrive at the thematic categories as presented in Table 5.2:

Table 5.2: AL for HE Thematic Item Pool

Scoping Review Code	Scoping Review Subthemes	Thematic Category	# items written
Knowledge and Understanding, Purposes of assessment, comprehension of standards	Knowledge base for assessment	Stable or Explicit Knowledge	5
Skills and Practice, integration in pedagogy	Disciplinary knowledge, SBA/SLO assessment, student assessment, knowledge of stakeholders	Flexible/Dynamic Knowledge	5
Metacognitions and reflection, response to external demands	Understanding promotes participation, metacognition,	Self-regulating Knowledge	5
Attitudes, Beliefs, Conceptions	Factors shaping conceptions,, beliefs, cognitions, limiting conceptions	Conceptions of Assessment	6
Fluency and Dialogue	Fluency with assessment language,	Fluency with Assessment Language	6
Community/collaboration, Contexts of assessment, Local Resources	Contexts, cultures of assessment, institutional characteristics,	Local Contexts	7
Decision-making	Compromise of competing demands in context, Using data, judgement of evidence	Decision-making	3
-	Development, Institutional support, Development in practice, need for development	Developing Assessment Literacy	5
Engagement and Experience, Demographic influences, Leadership, and advocacy	Professional role, experience and expertise, ownership	Engagement/Assessor Identity	7

The stem for these items was ‘Please indicate if you believe the following items to be true or false’. The response scale was dichotomous, labelled “True” or “False”. A dichotomous scale was selected to use these items to categorize respondents into groups based on typified beliefs about assessment in HE. For this purpose, it was desirable to prompt respondents to agree or disagree with statements that may have been polarizing. In addition, it was hoped that clusters of faculty may emerge who responded to opinion statements about assessment in similar ways based on their everyday demographic context.

In another section, items were written to reflect categories of HE professional tasks by consulting literature from the scoping review and brainstorming a variety of possible tasks for which HE faculty may need to involve assessment. Table 5.3 shows the proposed categorization of HE professional roles:

Table 5.3: HE Assessment Tasks Item Pool

Subscale name	# items
Teaching and Student Learning	13
Program Assessment	7
Faculty Service	16
Research	5
Statistical Dimensions of Measurement	5
Additional Roles	2

The stem for these items was “How much is assessment involved with each professional role or task below”? The response scale was Likert-style ranging from 1-5 with labels “Not at all”, “Slightly”, “Moderately”, “Very”, and “Extremely”. A Likert scale was selected to elicit respondent opinions about assessment involvement without forcing an extreme response. This type of response scale is appropriate when querying opinions about frequency of involvement because it accounts for nuance based on the everyday context of the respondent. Items were written to represent specific tasks necessary for each category of HE work. For example, the *Teaching/Student Learning* subscale contained items such as “assigning course grades”, “assessment of student learning”, “scoring classroom assessments”, and “student feedback”. One limitation of note is that it is very possible that additional purposes and functions exist for assessment in HE that were not accounted for by these items. To this end,

Following the intention to reflect as many thematic components and aspects of HE assessment as possible, this initial development strategy resulted in a survey that was long and thematically conflicted. There were 168 items across 3 sections encompassing proposed components of AL, HE professional tasks, and beliefs about AL in HE (see appendix H). This number of items reflected the three to four times the target survey length of ~50 items per recommendations by DeVellis (2017).

5.2 Broad Appeal and Item Review

The next phase of development was to explore the usefulness of the items to reflect aspects of AL in HE deemed significant to reviewers experienced with HE assessment and AL conceptualization. The purpose of such a review is to generate some evidence of content validity whereby experts in the content of the scale affirm the usefulness of the items to measure the target construct (DeVellis, 2017). For this thesis, content validity was somewhat problematic. The attempt to differentiate AL for HE from other kinds of AL meant a content universe had not yet been established, and items written to reflect scoping review themes were merely an attempt to represent a hypothesized construct. The scoping review attempted to describe content domains for each thematic component but may have been inaccurate.

Initially, expert review recruitment had been planned in a convenient way. Following a method used by Davies and Taras (2015), a review form for the survey would be solicited to assessment experts at an international conference in Spring 2020. The conference was cancelled

due to the COVID-19 pandemic, and the plan was adapted. Instead, review would be sought by soliciting reviewers who were academics working to conceptualize assessment-related constructs. An institutional IRB deemed an ethics application necessary for the expert review, and thus a proposal was submitted to the institutional ethics review board at the University of Saskatchewan (Appendix C).

Following approval, active researchers from various international institutions were contacted to provide feedback to reduce the number of items and confirm the usefulness of the instrument. The list of potential reviewers was formed by extracting names and contact information from the scoping review data. Authors were considered for expert review based on research activity in the particular domain of AL in HE. For example, authors who had attempted conceptualizations of AL were listed, as well as authors who were active in theorizing assessment-related concepts. Overall, 12 reviewers were contacted to participate, and five ultimately participated in the review. Those that declined to participate gave reasons of being on sabbatical, not having time, or simply did not respond. DeVellis (2017) had suggested using a panel of reviewers without offering a suggested number. For the purpose of content validation, the review proceeded with a panel of five reviewers. It is possible that the number of reviewers may not have been sufficient to represent all opinions about AL in HE, but it was expected that these perspectives would be helpful to explore the usefulness of the generated items thus far.

A review form was written that provided the theoretical framework for the survey content domains. The form instructed reviewers to read each item and rate each item from one (not useful) to three (very useful) as well as space to provide written feedback. The 1-3 item rating scale was selected to easily categorize potential items based on perceived usefulness without forcing respondents into a dichotomy. DeVellis (2017) suggested such a review may use a similar three-point scale to rate the relevance of each item to the construct being measured. As the target construct has not been sufficiently described in existing literature, a rating of “usefulness” was chosen as opposed to “relevance” because the target survey was hoped to be practical and exploratory. It was expected that feedback would vary based on reviewer bias: usefulness of a particular item was expected to vary based on the theoretical perspective eschewed by the contacted reviewers. For example, a dominant discourse in contemporary assessment literature is concerned with valuing assessment as far as it can be applied to improve student learning. Reviewers who conceptualized all assessment as primarily directed towards student learning assessment were expected to rate items pertaining to assessment for other uses in HE as ‘not useful’.

To account for the anticipated disagreement, it was determined to consider items with both unanimous and near-unanimous consensus on usefulness in the next iteration of the scale. To this end, the reviewer ratings for each item were summed and used to determine which items were considered most useful by reviewers. Items rated with greater sums would have been rated the most useful by the most reviewers.

As a target, the survey review aimed to reduce the existing survey to ~50 useful items with agreement from reviewers from varying perspectives on assessment. A scale of ~50 items

was targeted for several reasons. First, a longer survey was expected to fatigue potential respondents. As this survey was an attempt to generate empirical support for the thematic components extracted from the review, it was desirable to have a short, efficient survey that could easily probe the perspectives of HE faculty. Second, the scale was targeting three to five useful items for each of the eight major themes elicited from the scoping review, as well as \ representing agreement or disagreement with opinions about AL in HE. Third, it was anticipated that stress on HE faculty given the COVID-19 pandemic would impact response rates to the final survey. At the time of the review, the HE institution to be sampled from was closed to in-person work and many faculty were redeveloping resources for remote use. Thus, it was hoped that developing a shorter survey would limit the degree to which respondents discontinued their participation. Finally, reducing the survey from 168 items to ~50 would match the rate of item retention described by DeVellis (2017).

5.3 Reconceptualization and Redevelopment

In the next phase of development, a new version of the survey was compiled based on the reviewer scores and feedback comments. As expected, feedback on the scale made it clear that opinions differed starkly surrounding the construct of AL for HE. As well, reviewer consensus deemed the original version of the survey too bloated and complex for the purpose of the project. DeVellis (2017) cautions that advice from expert reviewers ought to be interpreted with caution as there is no accounting for the scale development experience of a given individual. For example, comments to reduce redundancy may constitute bad advice as redundancy informs internal consistency. The survey was reviewed by five experts familiar with assessment, but the views they reflect may not be reflective of the goal to survey faculty about AL in HE. As such, feedback from the reviewers was utilized in such a way that a survey reflective of the theoretical framework established by the scoping the review.

Feedback provided by reviewers varied. All reviewers commented on the number of items on the review form. However, it had not been explicated on the review form that many items had been brainstormed with the intent of discerning those most useful as per guidelines from DeVellis (2017). As such, it is possible that reviewers believed the review form to represent a more final version of the scale than intended. One reviewer offered critique that items unrelated to student learning assessment were entirely unrelated to the construct, but espoused full relevance of other conflicting elements. Comments of this nature had been expected based on diverse views about assessment present in literature, but it was useful to receive feedback suggesting that some content experts value particular functions assessment over other uses. Reviewers had also requested more context about the survey development framework, arguing that there was not sufficient evidence to suggest Kremmel and Hardings' (2020) approach could be adapted to AL for HE. Reviewers also made recommendations pertaining to scale choices and item formats. Finally, reviewers made recommendation regarding specific language conventions and clarity of wording for some items.

Overall, feedback from reviewers informed a decision to restructure the survey to be more straightforward. Using the original list of items, the survey items were recontextualized to focus on the themes that emerged from the scoping review analysis. As well, reviewer ratings of perceived item usefulness as well as qualitative comments were used to guide this process. Based on the summed reviewer scores, the original 168 items were reduced to 82 items deemed at least somewhat useful by all reviewers.

The item pool for section two had been written to reflect the component coding that occurred during the scoping review. Initially, this section was reduced from 66 items across 11 subscales to 28 items across 9 subscales. Specific expert feedback regarding these items included themes about conflicting structure and conceptual clarity as well as redundancy. Comments tended to reflect opinions about how assessment ought to be used in higher education. At this point in redevelopment, it was reflected that this item pool was redundant. A larger and more complex scale would be necessary to represent the entire range of codes from the scoping review. Focusing on these codes rather than the themes they contributed to created confusion within the content of the survey, and ultimately undermined the synthesis which had occurred in the thematic analysis. From the outset, it should have been assumed that these codes would be represented in items pertaining to the scoping review themes. To simplify the survey and redirect it towards its original purpose of the thesis, it was decided to remove this section and these items entirely. This was done to improve content validity as removal of these items would allow for a more targeted focus on the themes of the scoping review. However, it is acknowledged that removal of these items may result in a survey with less nuanced content by focusing on themes rather than components. Future investigations into AL in HE may consider investigating the nuance contained within each of these themes.

From the item pool for HE assessment tasks, 26 items across 3 subscales were chosen from 46 items across 6 subscales. The restructured task categories were Teaching/Student Learning, Program Assessment, and Faculty Service. Three expected categories of tasks were excluded based on a low number of items written and feedback from experts. For the few items for each of the removed task categories, there were conflicting opinions from reviewers regarding the usefulness of the items to represent aspects of HE work. Emphasis ought to be drawn to how feedback about these items: reviewers again gave opinion about assessment use in HE but noted that assessment may be involved in tasks depending on faculty roles and responsibilities. Further comments were made expressing concerns that “assessment is involved in these actions, but the assessment is notably different in each”. Though these comments have been expressed out of concern, they ultimately reflected the purpose of the survey to elicit opinions about *how* assessment is *differently* involved in these actions, which was in alignment with the guiding purpose of the survey.

From the item pool for AL in HE themes, 29 items over 7 subscales were selected from 49 items across 9 categories. These items were considered carefully due to the decision to remove the section of the survey relating to component codes. It was important that these items would accurately reflect codes and subthemes that emerged from the scoping review.

Considerations regarding the restructuring of these items are elaborated on in the following section.

After significant revision of the survey per the decision to restructure, 46 items remained. This survey reflected seven of the themes that emerged during the scoping review and a range of HE tasks across three categories. Each of the items that had been retained had near-unanimous agreement from the surveyed experts regarding usefulness to the survey. These items were further considered by the survey development team in relation to available literature. Items marked as important were rewritten and reorganized as described in the following section.

5.3.1 HE Professional Tasks

After rewriting, 18 items remained to reflect three groups of tasks in HE: Teaching/Student Learning, Program Assessment, and Faculty Service. Six items about Teaching and Student Learning were written to reflect various tasks involved with delivering learning to students at the course level in HE contexts. Two items about Program Assessment were retained to the use of assessments to evaluate academic programs and program outcomes. Further items for program assessment had been considered, but all these items were similar in content, and it seemed likely that they would contribute unnecessary redundancy. Ten items about Faculty Service were kept pertaining to tasks specific to HE faculty that may involve assessment, but not directly involve student learning at the course or program levels. There were more items retained for Faculty Service as it was not clear if faculty would agree with these items in a consistent way. These items are presented in Table 5.4 below:

Table 5.4: Higher Education Professional Task Items

Item	HE Task
1	Assessing student learning within courses (TSL)
2	Evaluating course materials such as textbooks or research artifacts (TSL)
3	Evaluating instruments for assessing student learning (TSL)
4	Critiquing assessment tools (TSL)
5	Student admissions decisions (TSL)
6	Engaging in feedback based on assessments (TSL)
7	Using assessments to evaluate program outcomes (PA)
8	Using assessments to evaluate academic programs (PA)
9	Advising students using different types of assessments (FS)
10	Evaluating faculty research artifacts (FS)
11	Evaluating theses/dissertations (FS)
12	Faculty evaluations based on peer assessment (FS)
13	Faculty evaluations based on student evaluations (FS)
14	Faculty hiring decisions (FS)
15	Faculty tenure/promotion decisions (FS)
16	Leading departmental assessment efforts (FS)
17	Participating in institutional assessment efforts (FS)
18	Research project management (FS)

Note: items categorized as: Teaching/Student Learning (TSL), Program Assessment (PA), and Faculty Service (FS).

These items were to be rated on a polychotomous scale including “Not at all”, “Occasionally”, “Sometimes”, “Often”, and “All the time”. Respondents could also describe a task as ‘Not Applicable’ to be scored as zero, creating a Likert-style scale with values from 0-5. Considerations for this scale are the same as previously described.

5.3.2 *HE in AL Themes*

After the review, 28 items were retained and rewritten to reflect seven thematic categories of AL for HE that emerged by synthesizing data from the scoping review (See Table 5.1). The items represent components of AL for HE that remained following the survey redevelopment. To improve dimensionality, items forming related groups were combined under expected labels. Six items were written representing both stable and fluid aspects of assessment theory and application. Three items represented metacognitive aspects of assessment including self-reflection and value derived from assessment data. Four items were written to query beliefs and conceptions of assessment as professionally useful, worthwhile, and connected to experience. Five items were written targeting each respondent’s identity as assessors, including

experience, engagement, and confidence. Five items were retained regarding fluency with assessment terminology including communication with other faculty and familiarity with assessment language. Three items were written to evoke aspects of assessment decision-making. Lastly, two items were written to query proclivity towards developing one's own assessment literacy. The full collection of these items are presented in Table 5.5 below:

Table 5.5: Thematic Component T/F Statements

Item	HE Task
18 (1)	Assessment processes result in information that is useful to me (M)
19 (2)	The same assessment principles underlie any assessment task (K)
20 (3)	Depending on who I am assessing, my interpretations could be different (K)
21 (4)	I can put my assessment knowledge into practice for any given task (K)
22 (5)	I have a practical working knowledge of assessment theory (K)
23 (6)	I recognize professional situations where I can apply my assessment knowledge (K)
24 (7)	Similar assessment concepts can be applied to different situations (K)
25 (8)	I reflect on my assessment knowledge and practices (M)
26 (9)	My opinion of assessment is shaped by my knowledge of assessment (M)
27 (10)	I regard assessment as useful in my profession (C)
28 (11)	My assessment efforts are generally worthwhile (C)
29 (12)	My views of assessment have been shaped by my experiences (C)
30 (13)	I am experienced at using assessment (I)
31 (14)	I consider myself to be an assessor (I)
32 (15)	I do not feel confident using assessment for tasks other than classroom assessment (I)
33 (16)	I do not need to practice assessment for making decisions (DM)
34 (17)	I feel confident about when and how to use assessment (I)
35 (18)	I clarify the meaning of the assessment concepts I use when discussing assessment with others (F)
36 (19)	I know how to communicate about assessment results and decisions (F)
37 (20)	I understand the concepts other people use when discussing assessment (F)
38 (21)	I use the term assessment in the same way as colleagues from other departments/colleges (F)
39 (22)	I use the term assessment in the same way as colleagues in my department/college (F)
40 (23)	I use assessment when I make professional decisions (DM)
41 (24)	In higher education, decision making is not possible without assessment skills (DM)
43 (25)	I generally do not need assessment skills to perform my tasks (C)
44 (26)	I participate in community learning activities regarding assessment (L)
45 (27)	I seek out opportunities to increase my assessment literacy (L)
46 (28)	I am engaged with assessment processes in my profession (I)

Note: Items categorized as Knowledge and Skills (K), Metacognitions (M), Conceptions and Beliefs (C), Assessor identity (I), Fluency (F), Decision-making (D), and Development and Learning (L).

As previously discussed, these items would be rated using a dichotomous scale labeled ‘True’ and ‘False’. This allowed for these items to be scored as groups of theorized components, but also for each statement to be used as a grouping variable for statistical analysis.

5.3.3 Open Response Items

Two open response items were included in the survey to further query faculty opinions on what tasks in HE involve assessment literacy. Participants were again provided with a definition of assessment in HE to prompt responses related to the guiding purpose of the thesis. The open response items are presented below in Table 5.6:

Table 5.6: Open Response Items

Item	Open Response items
47	Other than classroom assessment, how is assessment commonly used in higher education?
48	What do higher educators need to know and/or be able to do for these assessment tasks?

An open text field allowed participants to respond with any length and format. Responses were subject to a qualitative thematic analysis to complement statistical analyses of the survey data.

5.4 Deployment Strategy

All finalized survey items were imported to the online software SurveyMonkey for deployment (see full survey in appendix D). These included an informed consent statement, demographic items, 18 polytomous items representing three theorized groups of HE tasks, 28 dichotomous items representing seven theorized AL components, and two open response items querying faculty opinions on important aspects of assessment and assessment literacy in higher education. A stable link to the survey was created.

Deploying this version of the survey served a dual purpose of exploring how the survey items performed as well as gathering data to answer the research questions. To sample HE faculty, college deans at a large Canadian Prairie university were contacted and provided with the survey link to promote to their faculties. In addition, specific HE faculty with interests in assessment were directly contacted via email to request participation and recruitment of further participants. College deans were contacted after a two-week interval to follow up. The survey was live for five weeks at which point data were first downloaded.

5.5 Data Collection

The target population for the survey was HE faculty in the prairie provinces regardless of gender, length of service, area of study, or position held. A sample was targeted that would be of sufficient size to explore the underlying response patterns of the data set via a factor analysis. DeVellis (2017) explains that the stability of an emergent factor structure is at least somewhat a function of sample size, summarizing that a larger sample is generally better. DeVellis (2017)

conjectures that a sample of 150 is often used and cites other estimates ranging from 200 to 1000 participants. For the present scale, a sample of 200-300 faculty was the target. As will become evident through in the analysis, this was a difficult target to hit during this project. Significant constraints included time, cost, and access to the target population. As well, the survey was conducted at a time when HE faculty were largely working remotely during the COVID-19 pandemic. As such, the target population was under significant additional work pressure and stress due to the need to work with limited resources, adapt to working remotely, balance work and family responsibilities from home, and redevelop materials for remote use. It was expected that the additional work stress on HE faculty during this time likely limited willingness to respond to the survey. Ultimately, a final sample of only 55 responses was achieved, with three of those responses incomplete. This is significantly lower than recommended estimates for the target analyses as described above, and as such the results of the analyses should be considered with extreme caution.

The benefits of time, cost, and access to a difficult-to-reach population provided some justification for the sampling strategy. Ultimately, the awareness that the final sample constituted only 55 respondents recruited via convenience sampling informed the interpretation of the study. Subsequent analyses in the following chapter should be interpreted with the understanding that these data likely do not represent the opinions of the target population, and as such any conclusions about AL in HE should be considered rudimentary and exploratory. This limitation will be considered and discussed throughout the analysis.

With consideration to survey ethics, no participants were incentivized for their participation. All participants provided informed consent before completing the survey (see Appendix G), and no participants who responded to the survey were excluded from the analysis barring those who voluntarily exited the application prior to fully completing the survey.

5.6 Plan for Statistical Analysis

Statistical analyses for the study were conducted using R (R Core Team, 2018) and Jamovi (Jamovi, 2021). Descriptive statistics (e.g., frequencies, means, standard deviations, correlations) were used to describe characteristics of the sample (e.g., age categories, gender, position, and area of research), explore the dimensionality of the data, explore the data for significant correlations, and visualize emergent patterns. The outcomes of these analysis are discussed in the ‘Results’ chapter following discussion of the scoping review results.

CHAPTER SIX: SURVEY RESULTS AND ANALYSIS

The purpose of the survey study was to explore the separability of frequent HE tasks and to investigate potential correlations between HE tasks and theorized components of AL for HE. The HE task items were expected to form factors related to *Teaching/Student Learning*, *Faculty Service*, and *Program Assessment*. Further, such factors were expected to associate with individual statements related to AL in HE as well as theorized components of AL for HE: *Knowledge*, *Metacognition*, *Conceptions*, *Identity*, *Fluency*, *Decision-making*, and *Developing AL*. These correlations would mean that faculty who agree with different thematic statements about assessment believe assessment is used in HE in different ways. This would provide evidence to support the position that a unique conceptualization of AL for HE is necessary to account for how faculty use assessment in HE. The analysis was exploratory and as such no a priori hypotheses were determined. The following chapter describes the results from the descriptive and inferential statistical analyses of the survey data with respect to the research questions developed in the previous chapters.

With consideration to response bias, the present analysis is intentional in reporting limits to interpretation and non-significant results. The most pressing limitation was the sample size, which fell significantly short of the target n of 200 respondents. Despite the reduced sample size, analyses proceeded as planned with the caveat the interpretations may not be generalizable to the population (i.e. the institution from which the sample was drawn) nor HE faculty in broader contexts. Though possible extrapolations are limited, the analyses were considered useful as a very initial probe into faculty attitudes surrounding HE assessment, AL in HE, and HE assessment professionalization. The survey as described in the previous chapter ought not to be considered a final product, and the survey results described in the present chapter should be considered a baseline exploratory analysis from which future directions can be considered.

6.1 Sample Description

Probability sampling was not used for this project because the size of the target population was unknown. A significant constraint was faced in accessing the target population. At the time of the survey, HE faculty were largely working remotely during the COVID-19 pandemic. As such, the population was decentralized and only accessible via remote recruitment methods. In addition, the target population was under significant additional work pressure and stress due to the need to work with limited resources, adapt to working remotely, balance work and family responsibilities from home, and redevelop materials for remote use. It was expected that the additional work stress on HE faculty during this time likely limited willingness or availability to respond to the survey.

Access to HE faculty was reliant on academic deans at the target institution dispensing the survey to their faculty. As such, all participants were recruited via nonprobability convenience sampling and snowball sampling. Inherent limitations to convenience sampling include a lack of sufficient randomization and the strong possibility for a sample which is not representative of the entire population. The benefits of time, cost, and access to a difficult-to-

reach population provided some justification for the sampling strategy, but it is ultimately acknowledged that the population was not adequately sampled for the purposes of the present thesis. Consideration of the small size of the sample ($n=55$) is incorporated throughout the analysis.

The primary sampling methods were convenience sampling and snowball sampling at the target HE institution with the goal of recruiting at least 200 faculty from a diverse range of programs and with varying responsibilities in their everyday work. To access the target population, a list of academic deans was accessed. This strategy was selected because it was expected that HE faculty would be more likely to respond to a survey promoted by their departmental administration than a cold email from a graduate student. Deans were contacted via email with a description of the survey and a request to promote the survey to their faculty. Academic deans were reminded after two weeks, and those who had not yet replied were emailed again after a further week. Responding deans reported disseminating the survey to their faculty by email or by posting on departmental forums and message boards. In addition, snowball sampling was conducted by requesting that deans and faculty promote the survey to other HE faculty at the institution.

Final survey data were downloaded from SurveyMonkey on May 18th, 2021. At that point, the survey had been live for six weeks, and no more responses were forthcoming. The exported data showed that 77 surveys had been started. Further data collection was considered, but several challenges were considered. Most pressing, the sampling method had been described in the ethics application for the project. Further data collection would have been outside of these described parameters and thus would require further ethical approval which was not possible given constraints on the time and scope of the project. Further data could have been pursued by continuing to follow leads on academic deans who had not replied, but this considered to be antagonistic in light of the HE situation during COVID-19. At this point, it was determined to extract the data and proceed with the analysis under the consideration that the sample was much smaller than expected and below recommended thresholds for factor analysis (DeVellis, 2017). Despite the limitations of proceeding under these circumstances as described in the previous chapter, it was expected that interesting insights may be gleaned from an initial analysis of the data and a more representative sample could be sought in the future. At the very least, it was expected that qualitative data from the survey could be analyzed to explore if themes espoused by respondents were congruent with those elicited in the scoping review.

Data were conservatively organized and examined using Microsoft Excel to provide consistent variable names, exclude irrelevant variables (e.g., participant IP addresses, time to complete survey), and remove cases in which the survey was accessed but no data were gathered. After opening the survey, 22 participants did not continue past the demographics section and their responses were removed. After tidying the dataset, a total of 77 survey responses were reduced to a sample of 55 faculty responses for statistical analysis. As a further item of note, three respondents discontinued the survey after completing the Likert-style items, resulting in 52

responses for the dichotomous portion of the survey. Characteristics of the sample are described below.

The final set of respondents held various ranks at the institution, with most respondents holding ranks of assistant professor (20.00%), associate professor (30.91%), or full professor (32.73%). It was assumed that faculty of different rank would hold different responsibilities and invoke assessment differently for these responsibilities. The ranks of respondents are described in Table 6.1:

Table 6.1: Institutional rank of respondents

Rank at institution	<i>f</i>	%
Instructor	2	3.63%
Lecturer	2	3.63%
Assistant professor	11	20.00%
Associate professor	17	30.91%
Full professor	18	32.73%
Other	1	1.81%
No response	4	7.27%

As shown in Table 6.2, respondents identified as men or women in about equal proportions. Respondents had a range of experience as HE faculty as reported by their length of service, with most respondents (58.18%) having been employed as faculty from 10 to 25 years:

6.2: Respondent Characteristics

		<i>f</i>	%
Gender	Woman (including transgender women)	25	45.45%
	Man (including transgender men)	25	45.45%
	Prefer not to say	1	1.82%
	No response	4	7.27%
Years as HE faculty	<5	10	18.18%
	5 to <10	4	7.27%
	10 to <15	12	21.81%
	15 to <20	13	23.64%
	20 to <25	7	12.73%
	>25	5	9.09%
	No response	4	7.27%

6.2 Research Question One

What tasks are faculty performing in their everyday professional contexts that involve assessment?

In response to research question one, survey data were summarized to investigate the frequency of assessment tasks in HE and examined the means. The survey asked respondents to indicate how frequently assessment was involved with common tasks in higher education. The response scale was Likert-style with ordinal responses: “Not at all”, “Occasionally”, “Sometimes”, “Often”, and “All the time”. Respondents could also describe a task as ‘Not Applicable’ which is scored as zero to create a Likert scale with values from 0-5. The HE assessment tasks with means sorted from highest frequency to lowest, along with Cronbach’s alpha for each item, are presented in Table 6.3.

6.3: Mean frequencies of assessment involvement in HE tasks

Item	HE Task	Mean (SD)
1	Assessing student learning within courses	4.61 (0.63)
15	Faculty tenure/promotion decisions	3.93 (1.46)
13	Faculty evaluations based on student evaluations	3.81 (1.13)
11	Evaluating theses/dissertations	3.80 (1.57)
12	Faculty evaluations based on peer assessment	3.59 (1.14)
14	Faculty hiring decisions	3.59 (1.41)
6	Engaging in feedback based on assessments	3.56 (1.37)
2	Evaluating course materials such as textbooks or research artifacts	3.35 (1.03)
7	Using assessments to evaluate program outcomes	3.35 (1.17)
8	Using assessments to evaluate academic programs	3.31 (1.06)
3	Evaluating instruments for assessing student learning	3.19 (1.10)
10	Evaluating faculty research artifacts	3.13 (1.53)
9	Advising students using different types of assessments	2.96 (1.44)
4	Critiquing assessment tools	2.94 (1.14)
5	Student admissions decisions	2.89 (1.78)
16	Leading departmental assessment efforts	2.57 (1.35)
18	Research project management	2.52 (1.50)
17	Participating in institutional assessment efforts	2.46 (1.36)

The means for each of these items indicate each of these tasks in higher education involve assessment more frequently than “occasionally”. The most frequently reported assessment task *assessing student learning within courses* was the only task with a mean frequency of greater than ‘often’. This is reasonable given the documented frequency with which

assessment is used for these tasks in higher education, and specifically within AL literature. It is notable that the next most frequent task was faculty tenure/promotion decisions, a task that is prominently bound to HE settings and removed from student assessment purposes. Indeed, many of the tasks reported here as more frequently involving assessment are processes that could be considered bound to HE contexts such as faculty tenure/promotion decisions, evaluating theses/dissertations, and faculty hiring decisions. It is likely that higher frequencies for these tasks are related to the sampled population: HE faculty are likely to involve assessment in HE-related work. More importantly, most of the respondents (70.90%) were early- to mid-career academics with service lengths shorter than twenty years. Such a demographic has a vested concern with assessment for tenure/promotion in HE as they themselves are immersed in these processes. A conceptualization of AL for HE ought to be situated in the context of HE faculty, and these mean frequencies demonstrate how faculty place value on assessment processes with which they are involved or affected by. Such a finding underscores the importance of assessment for processes bound to HE contexts, but bear in mind that the sample size was small and may not be representative of the population.

Also of interest is examining the response data by standard deviation. Tasks with greater standard deviation suggest less agreement among respondents as to frequency with which assessment is involved with these tasks. As well as having the highest mean frequency, the task ‘assessing student learning within courses’ also deviated from the mean the least (0.63). This would suggest that respondents to this survey tend to agree that assessing student learning is one of the most frequent tasks for which they invoke their assessment expertise. Otherwise, most assessment tasks had standard deviations of greater than one suggesting the data is widely spread around the mean. One possible interpretation is that different faculty involve assessment with different tasks at different levels. Consider how the tasks with the highest standard deviations were student admissions decisions (1.78), evaluating theses and dissertations (1.57), and research project management (1.50). These wider deviations suggest a greater spread in how HE faculty consider assessment important for these tasks. Because the sample was drawn from a variety of programs and from faculty with a range of roles, it could be that departmental context has an important impact on how faculty frame assessment for their everyday work. Different disciplines frame assessment differently (e.g. Taras & Davies, 2017), and as such it is likely that contextual variation in assessment understanding was represented in these standard deviations.

The survey did collect demographic regarding home disciplines for each respondent, but respondents were distributed widely across programs. Given the limited sample size, the number of respondents for most programs was within single digits and it was determined not to analyze perceptions of assessment by program due to limitations to the interpretations of such results. Future studies achieving a larger sample may benefit from categorizing respondents by home discipline to determine which assessment tasks are perceived as most important for faculty operating in different academic contexts. Overall, the survey responses suggest that higher education faculty consider assessment as at least occasionally involved in a wide variety of

commonplace academic duties, supporting the position that a conceptualization of AL for HE ought to be developed to represent all facets of assessment in HE.

6.3 Research Question Two

How can common higher education assessment tasks be organized and labelled?

An exploratory factor analysis was selected to determine if there was an empirical basis for separability of assessment tasks in higher education and if so, how to describe those separations. A qualitative idea of what those tasks may have been emerged based on the previous scoping review, but there was no existing empirical research from to directly support these categories. Because of this, EFA was considered as an approach to explore the factor structure of the assessment tasks items. The EFA proceeded following established recommendations for such a procedure (e.g., DeVellis, 2017; Williams, Onsman, & Brown, 2010).

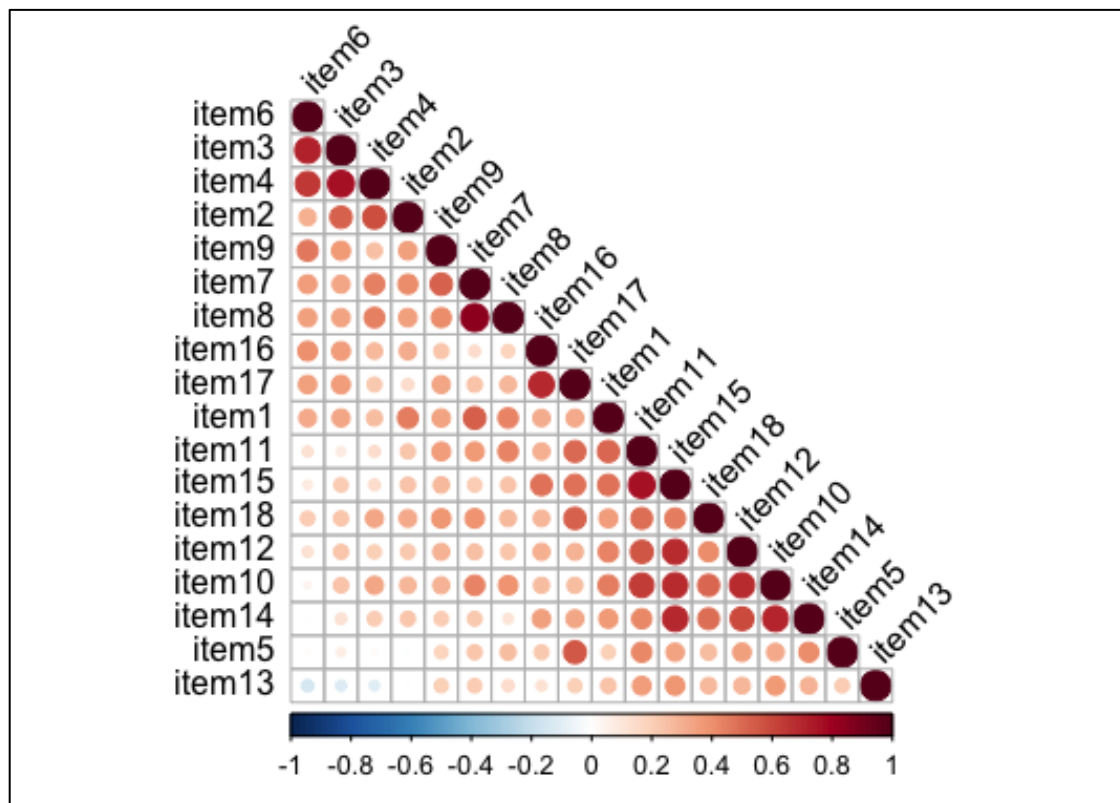
The EFA was conducted using Jamovi (Jamovi, 2021), a GUI for R (version 4.x), a language and environment for statistical computing and visualization. First, polychotomous data were reviewed and organized in a spreadsheet to ensure clerical, system-generated, and user errors were minimalized. Next, the data were inspected to ensure assumptions were met for factor analysis. Problematically, only a total of 55 faculty members had fully completed this portion of the survey. Though a conservative sample of ~200 had been initially targeted, this sample was even smaller than anticipated resulting in an item-to-participant ratio of 1:3. This ratio was line with older recommendations of a 1:3 ratio (Cattell, 1978), but far short of the 1:10 or 1:15 ratios contemporarily recommended to increase the power of the analysis (Bujang et al., 2012). This posed a significant issue for analysis: a sample this small likely does accurately represent the target population and may be insufficient to investigate the underlying factor structure of the data. As such, the following analysis should be interpreted with low confidence.

The decision to cautiously proceed with the EFA was based on several points. Williams, Onsman, and Brown (2010) summarize that for surveys with an item-to-participant ration below 1:5, the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy may be used to assess suitability of a sample for EFA. The KMO index for this sample indicated the sampling adequacy ($KMO = 0.758$) was acceptable as it was over 0.7, which is within the range of 0.5 to 1.0 recommended by Williams, Onsman, & Brown (2010). Next, it was reflected that the survey was exploratory in nature. From a highly conservative standpoint with regards to interpretation, it was decided to proceed this the analysis utilizing the existing smaller sample to examine if items seemed to group together as expected based on the theoretical framework established by the scoping review. These findings should not guide further scale development due to limitations with the sample, but yielded insights that may inform future projects to explore the factor structure of HE assessment tasks.

Regarding other assumptions for an EFA, Bartlett's test of sphericity was highly significant demonstrating that correlations between items were large enough for EFA ($X^2(153) = 704, p < 0.001$). As with the KMO index, Williams, Onsman, and Brown (2010) highlight that Bartlett's test of sphericity ought to be significant when conducting an EFA using a small

sample. A correlation table was generated and inspected which showed that items theorized to be related to one another seemed to be correlated ($> .30$), as demonstrated in the correlation heatmap in Figure 6.4:

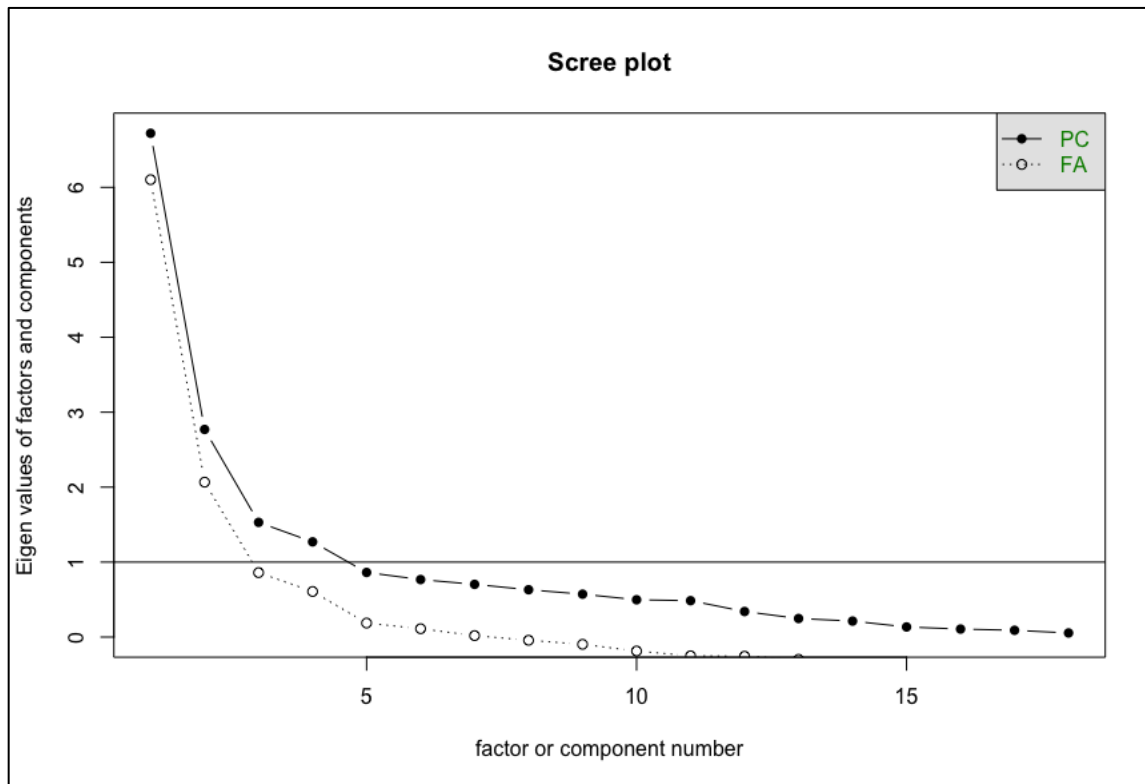
Figure 6.4: Heatmap of correlation strengths of survey items



This figure sorted items based on strength of correlation to help visualize groups of related items. Items 2, 3, 4, and 6 correlated more strongly with each other than with other items. These items had been written to represent teaching and student learning (TSL) tasks in HE. Notably, items 1 and 5 were also written to represent TSL tasks but responses to these items correlated more strongly with items 10 – 18 which were written to represent faculty service tasks. Finally, responses to items 7, 8, and 9, which were written to represent program assessment tasks, seemed to correlate with one another. All moderate or stronger correlations were positive. As well, moderate correlations between responses to items written to represent different kinds of HE tasks suggesting that any emergent factors of frequency of assessment in HE tasks may covary in similar ways.

Proceeding with the EFA, several strategies were utilized to estimate the number of factors. First, eigenvalues were calculated for possible factors. Two factors emerged with eigenvalues greater than 1, but inspection of a scree plot suggests a third factor may be a parallel analysis scree plot was created using R, as seen in Figure 6.5 below:

Figure 6.5: Scree plot of factor eigenvalues using different extraction solutions



This plot suggests two factors exist with eigenvalues greater than one. However, a third factor had an eigenvalue slightly below 1 and thus is worth considering, especially considering that it functions as the ‘elbow’ of the scree plot. Further to this, the scree plot for principal components (which partitions variance differently) suggests four components, but that the fourth may have been like the third. To investigate this further, a principal components analysis was performed, demonstrating that three components accounted for 61.2% of variance, a percentage considered acceptable when investigating social constructs. Given that items had been written to represent HE assessment tasks across three categories, it was determined to proceed with the factor analysis using three factors.

With three factors determined, a factor analysis was conducted using Principal Axis Factoring and a direct oblique rotation. An oblique rotation was appropriate given the expectation that items would correlate with one another because of interrelation among assessment tasks in higher education. Principal Axis Factoring was an appropriate factoring method because the analysis was probing the underlying factors of the response patterns rather than merely reducing the dimensionality of the data (e.g., Kremmel & Harding, 2020). Running the factor analysis with three factors resulted in an acceptable pattern matrix of loadings, displayed in Table 6.6 below:

Table 6.6: Initial Factor Loadings

	Factor			Uniqueness
	1	2	3	
item15	0.923			0.195
item14	0.798			0.402
item11	0.786			0.331
item10	0.755			0.327
item12	0.741			0.407
item17	0.576	0.389		0.417
item5	0.563			0.705
item13	0.561			0.659
item18	0.548			0.577
item1	0.459			0.420
item6		0.875		0.274
item3		0.869		0.230
item4		0.740		0.337
item16	0.455	0.493		0.442
item2		0.482		0.531
item9				0.612
item7			0.737	0.168
item8			0.619	0.322

Note. 'Principal axis factoring' extraction method was used in combination with an 'oblimin' rotation.

Two items, 16 and 17, cross-loaded onto two factors. A conservative approach to item retention was adopted by determining to iteratively remove items that cross-loaded on more than one factor at greater than .35. Note that with a larger sample with better probabilistic representation, it is possible that item loading may have been arranged entirely differently.

Item 16 (leading departmental assessment efforts) was removed first as it loaded more equivalently across factors 1 and 2 than item 17 (participating in institutional assessment efforts). After removing item 16, item 17 sufficiently loaded onto only factor one and the 3-factor solution remained viable. Eigenvalues for the final three-factor solution are shown in Table 6.7 below, explaining 60% of the variance in responses:

Table 6.7: Factor summary

Factor	SS Loadings	% of Variance	Cumulative %
1	4.85	28.5	28.5
2	2.87	16.9	45.4
3	2.52	14.9	60.3

As well, the full rotated pattern matrix after item removal is below in Table 6.8, as well as the factor correlations in Table 6.9:

6.8: Removed Factor Loadings

	Factor			Uniqueness
	1	2	3	
item15	0.936			0.1864
item14	0.847			0.3507
item12	0.772			0.3722
item10	0.756			0.3396
item11	0.724			0.3199
item18	0.533			0.5721
item5	0.503			0.7113
item13	0.495			0.6554
item17	0.480			0.6358
item1	0.415			0.4286
item3		0.952		0.1162
item4		0.809		0.2653
item6		0.754		0.3597
item2		0.489		0.5345
item7			0.935	0.0844
item8			0.840	0.2203
item9			0.378	0.6018

Note. 'Principal axis factoring' extraction method was used in combination with an 'oblimin' rotation

6.9: Factor Correlation Matrix

	1	2	3
1	—	0.353	0.485
2		—	0.514
3			—

Following the factor analysis, the items making up the identified factors were examined to infer a thematic label for each group of tasks. Factor 1 contained ten items representing professional processes context bound to HE settings including evaluating faculty peers, making advancement decisions, evaluating research products and processes, and assessing student learning. These items were assigned the label Faculty Service. Each of the four items in Factor 2 relate to evaluation of instruments and tools for assessment as well as using assessments for feedback. This factor was labelled Feedback and Evaluation. Finally, three items related to academic program assessment tasks such as evaluating program outcomes loaded onto Factor 3. This factor was labelled Program Assessment. Items for each factor were compiled into subscales, and Cronbach's alpha was calculated to indicate the reliability for each. Final labels, items, and reliability indices are presented in Table 6.10:

6.10: Revised Factor Summary

Factor	Assigned label	Item Numbers	α
1	Faculty Service	1, 5, 10, 11, 12, 13, 14, 15, 17, 18	.902
2	Feedback/Evaluation	2, 3, 4, 6	.872
3	Program Assessment	7, 8, 9	.825

Reliability estimates for the three-factor solution provide evidence of internal consistency for the three-factor solution, including Cronbach's alpha (.916) and McDonald's omega (.921).

To answer research question two, common HE tasks involving assessment as investigated in this survey can be organized into three empirically separable categories: Faculty Service, Feedback/Evaluation, and Program Assessment. For the most part, these categories of HE professional tasks are reflective of unique features of HE assessment as evoked in the scoping review. Items theorized as related to Teaching/Student Learning grouped as the unique factor feedback/evaluation with the notable exception of item 1 (assessing student learning within courses) and item 5 (student admissions decisions), which instead loaded onto the Faculty Service factor. This suggests separability regarding how frequently assessment is used in HE to assess student learning versus how it is used in feedback and evaluation processes at the course

level. Otherwise, items related to faculty service and program assessment loaded onto common factors in ways that were expected based on a priori theorization.

The factor analysis provides tentative support for the argument that a unique conceptualization of AL is necessary in HE to account for common assessment processes in HE. The factor structure that emerged from the survey data shows that assessment is used with differing frequency for distinct types of tasks. Further to this the removal an item that cross-loaded, as well as some cross-loading below the cut-off point of the analysis, suggests that additional factors exist which were not adequately represented in the present exploratory survey. Recall that the survey had a small sample and only a 1:3 item-to-participant ratio. Were the sample larger, it is possible that the underlying factor pattern of the data would be represented with stronger power, or else that items would load onto factors in entirely different arrangements. In such cases, it may be possible that the item would have loaded in a way that did not warrant removal. As a further example of the caution that must be taken interpreting these findings, the task “assessing student learning within courses” loaded the least strongly onto the *Faculty Service* factor as well as displayed cross-loading below the cut-off threshold. Despite this, it was rated by as the HE task in which assessment was most frequently involved. Given the primacy of student assessment in existing literature regarding AL itself, it seems likely that responses to a survey with more items reflecting specific HE student assessment processes may cluster to form an additional factor regarding student learning assessment. However, this survey was constructed with the intent of eliciting faculty perspective on HE assessment tasks without undue emphasis on student assessment, and these interpretations are taken with extreme caution given the small and non-comprehensive sample that was gathered.

6.4 Research Question Three

What relationships exist between common HE professional tasks and theorized components of AL for HE?

To examine this question correlations were computed between the separable factors of frequent HE assessment tasks and some theorized components necessary to AL for HE as elicited in the previous scoping review. First, item sums for extracted factors and theorized components were calculated to assign scores to each respondent. Items for HE assessment frequency items were grouped as factors emergent from the EFA: Faculty Service, Feedback/Evaluation, and Program Assessment. Next, components of AL for HE were scored into categories emergent from the scoping review and the importance determined by the survey expert review: knowledge, metacognition, conceptions, identity, fluency, decision-making, and developing AL. Each of these categories has been described in the earlier survey development chapter.

A Pearson correlation table was produced using these scores to estimate the relationships between theorized components of AL for HE and different types of assessment tasks. Normality tests and Q-Q plots suggested that the some of the data were not distributed normally, but the Pearson correlations proceeded as there were greater than 30 cases observed. Although it was expected that components of AL would be related to assessment frequency in HE tasks, a lack of

existing evidence of specific relationships meant that no a priori hypotheses were established regarding the direction of correlations. Further, these correlations ought to be interpreted with caution given the size and geographical limitations of the sample. Descriptive statistics for each variable and the full correlation matrix are presented in the tables below:

Table 6.11: Frequency Factor Descriptive Statistics

	Program Assessment	Feedback/Evaluation	Faculty Service
N	55	55	55
Mean	9.45	12.8	33.7
Median	10.0	13.0	36.0
Standard deviation	3.37	4.20	10.6
Minimum	0.00	0.00	0.00
Maximum	15.0	20.0	48.0

Table 6.12: AL Component Descriptive Statistics

	Knowledge and Skills	Metacognition	Attitudes, Beliefs, and Conceptions	Assessor Identity	Assessment Fluency	Decision- making	Development
N	53	53	52	52	52	52	52
Missing	2	2	3	3	3	3	3
Mean	3.98	2.66	3.75	3.58	2.69	2.54	1.02
Median	4.00	3.00	4.00	4.00	3.00	3.00	1.00
SD	1.42	0.706	0.682	1.58	1.38	0.779	0.804
Min	0.00	0.00	1.00	0.00	0.00	0.00	0.00
Max	6.00	3.00	4.00	5.00	5.00	3.00	2.00

Table 6.13: Correlation Matrix

	Faculty service	Feedback and evaluation	Program assessment	Knowledge	Metacognition	Conceptions	Identity	Fluency	Decision making	Learning
Faculty service	—									
Feedback and evaluation	0.400 **	—								
Program assessment	0.562 ***	0.604 ***	—							
Knowledge	0.157	0.331 *	0.295 *	—						
Metacognition	0.205	0.396 **	0.288 *	0.377 **	—					
Conceptions	0.419 **	0.248	0.189	0.332 *	0.505 ***	—				
Identity	0.411 **	0.191	0.291 *	0.225	0.042	0.337 *	—			
Fluency	0.260	0.130	0.201	0.155	0.269	0.229	0.137	—		
Decision making	0.374 **	0.319 *	0.249	0.205	0.272	0.590 ***	0.333 *	0.102	—	
Learning	0.030	0.059	0.172	0.069	0.252	0.223	0.362 **	0.253	0.202	—

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

The produced correlation matrix indicated significant relationships among frequent HE assessment tasks and AL components. As well, some significant relationships between the scored component groups of AL for HE; these are addressed in response to research question four below.

6.4.1 Relationships Between HE Assessment Tasks and AL Components

The frequency of assessment in Faculty Service tasks was moderately correlated with the AL components conceptions ($r = 0.42$, $p = 0.002$), identity ($r = 0.41$, $p = 0.002$), and decision-making ($r = 0.37$, $p = 0.006$). Faculty who more frequently involved assessment in Faculty Service tasks were also likely to conceive of assessment as positive or useful, identify themselves as assessors, and to consider assessment necessary in decision-making processes.

The frequency of assessment in Feedback/Evaluative tasks was moderately correlated with the AL components knowledge ($r = 0.33$, $p = 0.015$), metacognition ($r = 0.40$, $p = 0.003$), and decision-making ($r = 0.32$, $p = 0.021$). Thus, faculty who involve assessment with feedback and evaluation more frequently also tend to consider themselves to have sufficient theoretical and applied assessment knowledge. They tended to agree that they reflected on their own assessment knowledge and practice and considered assessment a major step in decision-making processes.

More frequent involvement in Program Assessment tasks was moderately correlated with knowledge ($r = 0.29$, $p = 0.032$), metacognition ($r = 0.29$, $p = 0.037$), and assessor identity ($r = 0.29$, $p = 0.037$). These correlations indicate that faculty who are involved with program assessment consider theoretical and applied knowledge and skills important to completing these tasks. They tend to reflect on their own assessment practice and consider themselves to be assessors.

In answer to the third research question, these correlations show that unique factors of HE professional tasks are positively related to different constellations of theorized components of AL for HE. Thus, HE faculty involve assessment differently in different types of tasks, and tend to agree with statements about assessment in different ways depending on the types of tasks they are involved in. As such, it seems plausible that elements of faculty AL may be invoked differently for assessment-related tasks in HE not directly supporting student assessment, countering existing conceptualizations of AL formulated in K-12 settings that focus on student assessment tasks.

6.5 Research Question Four

How is HE assessment and its involvement in HE professional work uniquely construed by HE faculty?

To investigate this research question, the correlation table produced above can be used to explore how faculty tended to respond to components about AL. As well t-tests were performed to explore more specifically how faculty involved assessment in HE tasks depending on how they responded to various component statements about AL.

6.5.1 Relationships Among AL Components

Significant relationships also emerged among theorized components of AL for HE. However, these correlations are interpreted cautiously given each component is the sum of statements theorized as grouped together. As such, these relationships were interpreted by examining the composite statements for each component. Knowledge was moderately correlated with metacognition ($r = 0.377$, $p = 0.005$) and conceptions ($r = 0.332$, $p = 0.016$). In other words, those who believe they have sufficient theoretical and applied assessment knowledge also tend to recognize the impact of their own cognitions on their assessment practice and consider assessment in a positive light. Additionally, metacognition was significantly correlated with conceptions ($r = 0.505$, $p < 0.001$). These three components were associated in a way that follows practical logic: those who know more about HE assessment also think about it and value it. However, review of the content domain of these items suggested overlap regarding item phrasing and it seems likely that they do not meaningfully differentiate the constructs they intend to represent. If these items had been scaled in a way to explore their factor structure (e.g., Likert-style) as opposed to being used as scored components and grouping variables, it seems likely they may load onto similar factors. Compounded with the limitations imposed by the sample, these interpretations are made with caution and not considered representative of the population.

Conceptions was also moderately correlated with identity ($r = 0.337$, $p = 0.014$) and decision-making ($r = 0.590$, $p < 0.001$). In other words those who conceive of assessment as useful also tend to consider themselves as experienced, engaged assessors, and consider assessment important to make decisions. Last, assessor identity was moderately correlated with fluency ($r = 0.333$, $p = 0.016$) and learning ($r = 0.362$, $p = 0.008$). In other words, those who consider themselves experienced assessors also tend to feel confident understanding and using assessment language and grammar, and tend to participate in opportunities to develop their AL. As above, these factors were expected to be related given the shared involvement and influence on HE assessment as emerged in the scoping review. Further, it is possible that the content domains for the theorized components represented other constructs than those intended.

6.5.2 Independent t Tests

To more thoroughly investigate how faculty uniquely construe of HE assessment, independent Student's t tests and Welch's t tests were performed on the HE task factors using each of the T/F components as grouping variables. This allows us to understand how the surveyed faculty involved assessment more frequently in the HE tasks factors based on how they responded to various statements about assessment. The t-tests were performed in Jamovi (Jamovi, 2021). Before each t-test the assumption of normality was investigated using Shapiro-Wilks' tests and quantile-quantile comparison plots. In some cases, the Shapiro-Wilk tests returned as significant, suggesting some non-normality; however, inspection of Q-Q plots revealed very slight violations of the assumption of normality. Given that the sample size was greater than 30, the t-tests proceeded. For cases where Levene's test of homoscedasticity returned a significant statistic, Welch's T was reported instead of Student's T. Welch's T is a

more conservative test that does not assume equal variances. Finally, for some items the majority of respondents fell into one group. For example, on item 23 there were 49 respondents in the ‘True’ group and 3 in the ‘False’ group. Statistics from these tests were not reported.

Statistics from these tests that were significant at an alpha level of 5% are reported in Table 6.14:

6.14: AL in HE Analysis

Item	Program Assessment	Feedback/Evaluation	Faculty Service
I can put my assessment knowledge into practice for any given task	Significant (T group) $t(51) = -2.07, p = 0.043$	Significant (T group) $t(51) = -2.42, p = 0.043$	Significant (T group) $t(51) = -2.64, p = 0.011$
I recognize professional situations where I can apply my assessment knowledge	Significant (T group) $t(51) = -2.74, p = 0.008$	Significant (T group) $t(51) = -2.67, p = 0.010$	Nonsignificant
My assessment efforts are generally worthwhile	Nonsignificant	Nonsignificant	Significant (T group) $t(51) = -2.93, p = 0.005$
I am experienced at using assessment	Significant (T group) $t(51) = -2.36, p = 0.022$	Nonsignificant	Significant (T group) * $t(51) = -3.60, p = 0.002$
I do not feel confident using assessment for tasks other than classroom assessment	Nonsignificant	Nonsignificant	Significant (F group) $t(51) = 2.482, p = 0.016$
I feel confident about when and how to use assessment	Nonsignificant	Significant (T group) $t(51) = -2.38, p = 0.021$	Nonsignificant
I know how to communicate about assessment results and decisions	Significant (T group) $t(50) = -2.28, p = 0.027$	Nonsignificant	Significant (T group) $t(50) = -2.62, p = 0.012$
In higher education, decision making is not possible without assessment skills	Significant (T group) $t(50) = -3.19, p = 0.002$	Significant (T group) $t(50) = -3.04, p = 0.004$	Significant (T group) $t(50) = -2.69, p = 0.010$

Note: For statistics marked with “*” Levene’s test was significant, and Welch’s T is reported.

These results are helpful to understand how faculty who have more involvement in particular assessment tasks respond to particular statements about assessment. Overall, faculty involved in different types of assessment tasks responded in different ways to different statements.

Faculty who agreed with “I can put my assessment knowledge into practice for any given task” perceived assessment as more involved in Program Assessment ($t(51) = -2.07, p = 0.043$), Feedback/Evaluation ($t(51) = -2.42, p = 0.043$), and Faculty Service ($t(51) = -2.64, p = 0.011$). Similarly, faculty who agreed that “decision-making is not possible without assessment skills”

had significantly higher mean frequencies of assessment involvement in Program Assessment ($t(50) = -3.19, p = 0.002$), Feedback/Evaluation ($t(50) = -3.04, p = 0.004$), and Faculty Service ($t(50) = -2.69, p = 0.010$). Thus, faculty who feel capable of applying their assessment knowledge and considered it necessary for making decisions were more likely to involve assessment more frequently in all types of HE assessment tasks.

Those who agreed with “I recognize professional situations where I can apply my assessment knowledge” rated significantly more frequent assessment involvement in Program Assessment ($t(51) = -2.74, p = 0.008$) and Feedback and Evaluation ($t(51) = -2.67, p = 0.010$), but not Faculty Service. This result suggests that faculty who discern opportunities to apply assessment knowledge involve assessment more frequently in more specific types of tasks. It may be possible that the Faculty Service factor represents tasks where the frequency of assessment involvement is explicit, or else that these types of tasks are routine for HE faculty and do not necessitate explicit professional application of assessment knowledge.

Indeed, faculty agreeing that their “assessment efforts are generally worthwhile” were significantly more likely to involve assessment in Faculty Service tasks ($t(51) = -2.93, p = 0.005$) but not in other tasks. This suggests that those who feel their assessment effort yields worthwhile results tend to involve assessment more frequently in tasks typical of HE professional work. Taken with the previous result, it may be possible that faculty who perceive their efforts as worthwhile typically involve assessment more frequently in Faculty Service tasks and perceive opportunities to involve assessment in other kinds of tasks.

Faculty who agree that they are “experienced at using assessment” involve assessment significantly more frequently in Program Assessment ($t(51) = -2.36, p = 0.022$) and Faculty Service (Welches’ $t(51) = -3.60, p = 0.002$) tasks, but not Feedback/Evaluation tasks. Similarly, faculty agreeing that they “know how to communicate about assessment results and decisions” had significantly higher mean frequencies of assessment involvement in Program Assessment ($t(50) = -2.28, p = 0.027$) and Faculty Service ($t(50) = -2.62, p = 0.012$) tasks. These results may indicate that as faculty gain experience using assessment they are more willing to involve it in tasks outside of the feedback and evaluation tasks represented in this survey which included tasks expected to be performed by faculty involved in teaching and student learning. It may also be the case that more experienced assessors have attained higher ranks within their institution and have a more diverse profile of professional responsibilities.

Faculty who agreed with the statement “I do not feel confident using assessment for tasks other than classroom assessment” were significantly less likely to involve assessment in Faculty Service ($t(51) = 2.482, p = 0.016$). As for the inverse, it seems reasonable that faculty who do feel confident using assessment for tasks other than classroom assessment would be more likely to do so more frequently. Interestingly, faculty who agree that they “feel confident about when and how to use assessment” were significantly more likely to involve assessment in Feedback/Evaluation tasks ($t(51) = -2.38, p = 0.021$), but not other kinds of tasks. Based on these tests, it could be that confident faculty *perceive* the assessment involved in Feedback/Evaluation tasks to a greater degree than non-confident faculty.

Overall, these t tests illuminated how faculty attitudes about specific aspects of assessment are related to how they involve assessment in different types of HE tasks.

6.6 Thematic Analysis of Qualitative Items

To further support these ideas, qualitative survey data can also be explored. To this end the data were thematically interpreted from the two qualitative questions: “Other than classroom assessment, how is assessment commonly used in higher education?”, and “What do higher educators need to know and/or be able to do for these assessment tasks?”. For each of these items, each response was read and coded based on content (Braun & Clarke, 2006). For example, the comment “acceptance of students to programs” was coded as “admissions”. Similar codes were iteratively and progressively grouped together until thematic patterns emerged from the data. For example, the following codes were grouped based on thematic similarity: “Hiring, tenure, and promotion processes”; “faculty evaluations”; “hiring of staff”; and “assessing awards”. Each group of codes was assigned a thematic label reflecting elements and features of HE assessment as described by the responding faculty. In the above example, the related codes were assigned the label “Faculty Management/Decision making”.

6.6.2 Open Ended Item One

For the first item (“Other than classroom assessment, how is assessment commonly used in higher education?”), this process resulted in nine thematic labels containing various subthemes. These themes are described below in order of frequency. For this section, excerpts from respondent comments are provided in quotations without citations.

The most common theme was labelled faculty management & decision-making and was referenced 38 times across three subthemes: Hiring, tenure, and promotion processes (22), Faculty teaching evaluations (9), and assessing awards (7). Hiring, tenure, and promotion processes emerged as particularly important to faculty responding to this survey. This subtheme described specific processes such as assessment of faculty teaching, research productivity, and efficacy with administrative tasks. One commenter elaborated how such decisions incorporate “student and peer feedback on teaching, assessments of research quality and quantity, and assessments of administrative tasks”. Similarly, the subtheme of faculty evaluation reflected appraisal and review of faculty teaching and performance, while the subtheme assessing awards demonstrated faculty interest in assessment of merit cases and awards including those with a financial incentive. It is reasonable that faculty identified these processes as areas where assessment is invoked: most faculty participate in such processes as both assessor and assessee, and such evaluations occur with relative frequency. One commenter explained:

“As a professor, I am submitted to multiple teaching assessments each year, including three peer reviews of my teaching, and multiple student learning evaluations [...] these assessments are central to the tenure and promotion process.”

Further, such processes are tied to career progression paths and financial benefits for faculty. It is thus notable that faculty would identify such tasks in the context of AL. Perhaps because of the direct impact in terms of workload and incentives, these processes may hold unique significance for HE faculty who frequently participate and stand to benefit from them.

The second theme was labelled student decision-making and was referenced 30 times across 3 subthemes. Graduate/doctoral student management (15) includes processes like approving thesis/dissertation proposals, providing feedback, and assessment of completed research products. This subtheme represented summative elements of “determining progress in thesis-based programs” as well as formative “evaluation of graduate student research skills and other soft skills”. A related decision-making subtheme was admissions (10) pertaining to “admission”, “recruitment”, and “acceptance of students to programs”. Responding faculty particularly identified the selection of graduate/postdoc students; it is possible that most teaching faculty are more involved in these selections than they are in undergraduate admission processes. Last, despite a prompt to identify uses for assessment outside of assessing learning, managing student learning (5) was identified in relation to discipline-specific student assessment opportunities including experiential learning, clinical settings, and authentic assessments.

The theme program assessment was referenced 27 times across three subthemes. Program evaluation (17) processes included systematic program review, assessment, and modification, as well as “seeking feasibility data on launching new research and teaching directions and programs”. Many of these references were typified and nonspecific suggesting terms like ‘program assessment’ may be more formal and stereotyped, or that faculty understand assessment as involved in the process but are familiar with the process by their participation rather than possessing an overarching comprehension of the process. Curriculum development (6) emerged as distinct subtheme of program assessment involving tasks such as course design, curriculum review, and program renewal. Finally, accreditation purposes (4) were identified as a distinct engine for program assessment efforts in HE, underscoring the idea that some faculty in HE are more familiar with program assessment as a formal process in which they are participants without ownership.

Resource management was referenced 17 times across 3 subthemes: managing research funding (9), budgeting/allocating resources (5), and non-faculty hiring (3). The theme of resource management primarily invoked assessment processes in decision-making processes appearing more administrative in nature. Managing research funding reportedly involved assessment in such tasks as reviewing grants, awarding internal funding, and reporting to funding agencies. Budgeting and allocating resources included selection and purchase of equipment as well as “activity-based budgeting”. Finally, three examples reflected non-faculty hiring including administrators, summer students, and departmental staff.

The theme disciplinary faculty service had 26 references across 5 subthemes. Predominantly, scholarship (12) involved assessment throughout the academic writing processes including research studies, manuscript review, and peer review of submitted and published scholarship. Discipline-specific assessment (6) centered on selecting “appropriate assessment for

certain tasks” bound to certain disciplines, such as legal consultation, “diagnostic and clinical duties”, and establishing and reviewing disciplinary “professional practices and standards”. One respondent clarified this subtheme insightfully by highlighting how assessment is understood and utilized within their profession:

“We are unique. We are a professional college so assessment can mean assessment as a teacher/educator or a professional veterinarian. These are not the same assessment.”

Comments such as these highlight the distinction between aspects of HE AL necessary for student assessment versus other assessment purposes in HE settings and underscores that AL as it has been defined in K-12 contexts cannot be seamlessly ported to more complex HE contexts.

Next, the subtheme scientific processes (4) described how “all forms of science use assessment” including research design, data collection, assessment of data quality, and evaluation of research results. This follows rationally as one respondent elaborated how “gathering and interpreting evidence forms the basis for proving or disproving a hypothesis”, underscoring the ubiquitous importance of AL for any faculty involved in producing or utilizing research in HE contexts. Finally, lesser represented subthemes included: committee work (3) such as “participating in review committees for collegial processes”; and ethics (1) through “examining issues of equity, diversity, and inclusion”.

The theme metacognition had eight references reflecting subthemes of self-improvement (4) and personal awareness (4). Self-improvement reflected using assessment or assessment data to improve one’s own assessment practice including “informal peer assessment”, coaching, and student feedback. The subtheme personal awareness contained reflective comments made by several respondent faculty used to qualify their responses, such as: “I can only speak to my own disciplinary knowledge”. These few but important comments identify an important mediating metacognitive step of critical reflection about how AL is applied to specific tasks interpretively and subjectively.

Six comments referenced administrative decision making with specific subthemes about central administration (3) and operation planning (3). Central administration involved measuring progress on institutional goals and “reporting to stakeholders such as government”. Identification of these tasks reflects their importance in HE, but the distance from which faculty responding to this questionnaire have from such tasks. Similarly, business and operation planning subthemes referred to using assessment for evaluation of strategic plans as well as old and new policy.

Finally, four comments reflected themes of miscellaneous assessment opinions such as assessment ubiquity (2) and assessment misuse (2). Assessment ubiquity described how assessment is important in a diverse range of processes across HE contexts, with comments expounding how assessment “informs virtually every process or should” for “most activities on campus”. Conversely, two comments about assessment misuse identified frustrations with how HE assessment is enacted by peers who “do not have a good understanding of assessment”.

Overall, the thematic analysis of the first open-ended item provided valuable insight as to how assessment is specifically invoked in HE contexts. The qualitative survey data provided useful elaboration on how HE faculty view assessment as used within their own contexts. These data are particularly useful for the purpose of understanding how an AL framework may be developed for HE contexts. This data provides useful structure to future work investigating what contextual assessment knowledge and skills are necessary for efficacious assessment work in HE settings.

6.6.2 Open Ended Item Two

For the second item (What do higher educators need to know and/or be able to do for these assessment tasks?), this process resulted in 6 thematic labels containing various subthemes. The themes and subthemes are described below in order of frequency.

With 21 references, the most common theme was knowledge of assessment theory which reflected components of theoretical assessment understanding deemed necessary by respondents for efficacious assessment use in HE contexts. The first of two subthemes, assessment theory (13), labelled specific aspects of assessment knowledge which might be included in a conceptualization of AL for HE. These typically reflected basic concepts of assessment and psychometrics such as reliability, validity, fairness, and appropriate generalization; however, the extent of necessary knowledge ranged from “basic understanding” or “foundational knowledge” to comprehension of the “entire body of assessment theory and practice”. The second subtheme was labelled purpose of assessment (8) reflecting the “meaning and purpose of particular assessment activity”. These referents centered on the necessity of outlining the intent of a particular assessment activity in order to select specific and appropriate assessment practices. Foundational understanding of assessment theory intuitively underlies the selection of assessment techniques to fit a particular purpose; it seems possible that the integration of such knowledge and selection may be necessary for AL in HE.

The next most common theme, knowledge of assessment application, was referenced 15 times across four subthemes: Standards/benchmarks (8), assessment practice (5), assessment in science (3), and student learning assessment (1). These comments follow logically from the previous theme: a base of assessment knowledge and purpose is expanded on upon by practicing assessment. The subtheme standards/benchmarks was referenced eight times in reference to utilizing benchmarks, standards, and criteria in various applications including establishing fair and valid criteria, becoming familiar with existing standards and benchmarks, and what “kinds of evidence support claims that standards/criteria are met or not”. More generally, five comments with the subtheme assessment application spoke to the necessity of “being able to [create] and assessment that matches the goals of the task” – that is, the need to practice assessment in accordance with foundational knowledge and established purposes. This includes using appropriate tools, selecting assessment methods to fit each task, and comprehension of “the knowledge, skills, attitudes, or competencies being assessed”. To a degree, this necessitates maintained grasp on the broader picture of assessment use and resistance to being bogged down

by the minutia of process. Last, three comments referred to assessment in science, including using assessment to “gather and analyze empirical data and evidence” in research contexts; and one comment referred to student learning assessment skills emphasizing that all higher educators, regardless of their specific specialty area, are teachers”. This reinforces the existing notion that assessment literacy is necessary for student learning components of HE professional responsibility in addition to other functions.

The theme development was referenced 12 times across four subthemes: assessment skill practice (4), institutional processes (4), resources (2), and student feedback (2). This theme reflected administrative resources that faculty feel are necessary for overall improvement of practice of assessment in HE. Comments referencing assessment skill practice advocated for chances to practice different aspects of assessment including “building and evaluating” assessments. As well, some claimed the need for “opportunities to apply different models of assessment” which reflects a desire for space to experiment outside of institutionally sanctioned assessment processes. Similarly, four comments were marked with the subtheme of institutional processes referring the necessity of top-down institutional support for assessment as well as bottom-up “sharing of ideas, successes, and failures with colleagues”. Two additional comments explicitly requested institutional resources to learn about assessment, including extra “time to think and plan”. Finally, two comments highlighted openness to student feedback “regardless of [research] area”, re-emphasizing the professional faculty role as an educator among other responsibilities.

Next, metacognitive skills were referenced 12 times over three subthemes: critical thinking (6), limiting bias (4), and personal role in assessment (2). Critical thinking was referenced as “a key component of assessment” in a variety of uses including reading and using data appropriately, as well as “considering as many factors as possible when assessing [in order to] prioritize them”. One insightful comment explained how “...higher educators need to be able to read between the lines – to hear what isn’t being said, see who isn’t at the table, and understand whose voice is not being heard and why”. From this view, a component of AL is an understanding of the parameters of an assessment to make sound judgements regarding applicability of data. This orientation to assessment fairness connects to another subtheme, limiting bias, which was assigned to comments about maintaining a “decentralized viewpoint” in order to “recognize and address bias”. AL in HE contributes to such fairness; one who is assessment literate can orient their assessment knowledge and understand to develop “a framework for making assessments ... that is as objective as possible”. The final subtheme was self-reflection, which espoused faculty understanding of “their own role” in order to “...assess ourselves and our peers”.

Context was referenced eight times in two subthemes. Conventions of the discipline (4) was referred four times in relation to understanding “differences between different colleges ... for assessment tasks to remain relevant”; for example, some commenters referred to discipline-specific assessment processes such as within dentistry, legal, or medical contexts. Indeed, context of practice (4) was a subtheme used to label comments about “application of theory and

knowledge under different assessment scenarios”. Context seems to be understood as an intersubjective situational variable mediating how one’s own assessment knowledge and skills are expressed within an assessment environment. This is expressed as a situational awareness of factors that may impact assessment processes and outcomes including the personal context of assesseees, the dependence of a particular assessment task on environmental variables, or disciplinary factors impacting assessments. As such, context as a subtheme is important for AL in HE given the high level of subjectivity among assessment uses and scenarios. Personal sensitivity to all such contextual variables seems as necessary as understanding of assessment theory, purposes, or application in context.

Last, seven references were made to deficits in HE assessment reflecting perceived deficits in training, understanding of purpose, participation, and collaboration. Most commonly, survey respondents identified a lack of training as a barrier to effective assessment practice in HE; one respondent bluntly summarized that assessment “is not something we are formally trained in by and large, and thus we may not be doing it in the optimal ways”. Further comments expressed judgement about assessment deficits such as how faculty “should be trained how to use assessment properly, but most are not”. Other specific comments reflected specific areas for improvement. One participant felt some constructs assessed in HE are not thoroughly operationalized, such as “originality, preparedness to undertake research, or intellectual capacity”. Another faculty member felt frustrated that their “colleagues specifically do not use assessment as much as they could to make departmental decisions”. A third participant seemed to agree, noting that “when they suggest assessments to help us gather empirical evidence to make decisions, I just get shut down”. Each of these deficits suggests a direction for growth and development of AL for HE.

6.7 Summary of Survey Findings

To conclude this chapter, a summary of results for each survey research question is reported. First, survey responses suggested that HE faculty consider assessment as at least occasionally, and typically sometimes, involved in a wide variety of commonplace academic duties, supporting the position that a conceptualization of AL for HE ought to be developed to represent facets of assessment additional to student learning assessment. Next, common HE tasks involving assessment as investigated in this survey can be organized into three empirically separable categories: faculty service, feedback/evaluation, and program assessment.

Significant Pearson correlations were found between assessment involvement in HE tasks factors and theorized AL for HE components. Faculty Service tasks were moderately correlated with conceptions ($r = 0.42$, $p = 0.002$), identity ($r = 0.41$, $p = 0.002$), and decision-making ($r = 0.37$, $p = 0.006$). Feedback/Evaluative tasks were moderately correlated with knowledge ($r = 0.33$, $p = 0.015$), metacognition ($r = 0.40$, $p = 0.003$), and decision-making ($r = 0.32$, $p = 0.021$). Program Assessment tasks were moderately correlated with knowledge ($r = 0.29$, $p = 0.032$), metacognition ($r = 0.29$, $p = 0.037$), and assessor identity ($r = 0.29$, $p = 0.037$).

Significant Pearson correlations were also found among theorized components of AL for HE. Knowledge was moderately correlated with metacognition ($r = 0.377$, $p = 0.005$) and conceptions ($r = 0.332$, $p = 0.016$). Metacognition was significantly correlated with conceptions ($r = 0.505$, $p < 0.001$). Conceptions was also moderately correlated with identity ($r = 0.337$, $p = 0.014$) and decision-making ($r = 0.590$, $p < 0.001$). Assessor identity was moderately correlated with fluency ($r = 0.333$, $p = 0.016$) and learning ($r = 0.362$, $p = 0.008$).

Independent t-tests flagged significant differences among assessment involve in different task factors using T/F statements as grouping variables. Faculty who agreed with “I can put my assessment knowledge into practice for any given task” perceived assessment as more involved in Program Assessment ($t(51) = -2.07$, $p = 0.043$), Feedback/Evaluation ($t(51) = -2.42$, $p = 0.043$), and Faculty Service ($t(51) = -2.64$, $p = 0.011$). Faculty who agreed that “decision-making is not possible without assessment skills” had significantly higher mean frequencies of assessment involvement in Program Assessment ($t(50) = -3.19$, $p = 0.002$), Feedback/Evaluation ($t(50) = -3.04$, $p = 0.004$), and Faculty Service ($t(50) = -2.69$, $p = 0.010$). Faculty who agreed with “I recognize professional situations where I can apply my assessment knowledge” rated significantly more frequent assessment involvement in Program Assessment ($t(51) = -2.74$, $p = 0.008$) and Feedback and Evaluation ($t(51) = -2.67$, $p = 0.010$), but not Faculty Service. Faculty agreeing that their “assessment efforts are generally worthwhile” were significantly more likely to involve assessment in Faculty Service tasks ($t(51) = -2.93$, $p = 0.005$) but no other tasks. Faculty who agree that they are “experienced at using assessment” involve assessment significantly more frequently in Program Assessment ($t(51) = -2.36$, $p = 0.022$) and Faculty Service (Welches’ $t(51) = -3.60$, $p = 0.002$) tasks, but not Feedback/Evaluation tasks. Faculty agreeing that they “know how to communicate about assessment results and decisions” had significantly higher mean frequencies of assessment involvement in Program Assessment ($t(50) = -2.28$, $p = 0.027$) and Faculty Service ($t(50) = -2.62$, $p = 0.012$) tasks. Faculty who agreed with the statement “I do not feel confident using assessment for tasks other than classroom assessment” were significantly less likely to involve assessment in Faculty Service ($t(51) = 2.482$, $p = 0.016$). Lastly, faculty who agree that they “feel confident about when and how to use assessment” were significantly more likely to involve assessment in Feedback/Evaluation tasks ($t(51) = -2.38$, $p = 0.021$).

In the next chapter, these statistics are further discussed in the context of the thesis as a whole. They are related to findings from the scoping review and interpreted in light of qualitative analyses, existing literature, and limitations to the study.

CHAPTER SEVEN: DISCUSSION AND CONCLUSION

This chapter interprets and discusses the findings of the survey study, compares them to scoping review findings, and explores their implications in light of contemporary HE assessment literature and practice. It will begin with discussion of the results of each research question with reference to the literature discussed in Chapters Two and Three. Next, a brief discussion will integrate the results of each research question with each other as well as with HE assessment literature in general. Finally, limitations to the research will be discussed, future directions for research will be explored, and a conclusion to the thesis will be drawn. Overall, it is intended for this discussion to gird future empirical work to conceptualize AL for HE contexts.

Educational assessment is irrevocably ingrained in contemporary Canadian higher educational models. It is an increasingly imperative tool for educational accountability and improvement as external stakeholders increasingly call for evidence that postsecondary education efficaciously improves outcomes for learners (Ewell, 2009; Huber & Skedsmo, 2016; Sarrico et al., 2010). As assessment becomes more widely utilized, it is important that stakeholders operating in HE settings demonstrate sufficient AL to utilize assessment for an expanding range of tasks such as: assessing student learning, assessing program effectiveness, interpreting and applying assessment data correctly and ethically, and making data-informed decisions (Arreola, Theall, & Aleamoni, 2003; Medland, 2019; Schoepp & Tezcan-Unal, 2017). As a professional skill, AL in HE involves clear comprehension of the theory and principles that underlie educational assessment to such a degree that they may be appropriately implemented in practice, as well as discussed with other stakeholders in assessment processes (Davies & Taras, 2018; Medland, 2019; Popham, 2011). This thesis argues that AL for HE must be uniquely conceptualized because HE contexts invoke contextually situated assessment knowledge, understanding, and skills in ways that have not been accounted for by traditional HE literature (Medland, 2019).

A scoping review investigated this problem and identified a that existing AL conceptualizations did not sufficiently account for all professional responsibilities of higher educators. Adopting a systems approach to establish a contextual theoretical frame, the review espoused how AL can be viewed as a micro-level influence on practice (Fulmer, Lee, & Tan, 2015). In this view, AL is an individual level contextual factor, alongside other contextual factors, having an impact on assessment practice. Fulmer, Lee, & Tan (2013) framed AL in this way using the classroom setting, but this theoretical framing is appropriate to situate AL within other settings, purposes, and professional tasks. For example, rather than the statement, “AL is a micro-level influence on student learning assessment practice”, other purposes or roles might be substituted in the place of student learning: “AL is a micro-level influence on program assessment practice, or language assessment, or professional assessment”.

At the outset, the aim of this thesis was to conceptualize what a contextualized AL for HE may be composed of, and to explore how that AL for HE may diverge from AL for school-based settings. The scoping review demonstrated that to do this, a better picture of assessment involvement in HE professional work was needed. To explore this, HE faculty were surveyed to

gain an idea of how HE assessment is involved in professional work including that assessment which is not attached to student learning assessment.

7.1 Research Question One

What tasks are faculty performing in their everyday professional contexts that involve assessment?

It was found that assessment was perceived as involved at least occasionally (and typically more frequently) in a diverse selection of tasks across diverse HE contexts, and as particularly involved in processes with direct impacts on the faculty themselves. Assessment was most frequently involved in “assessing student learning within courses”, “faculty tenure and promotion decisions”, and “faculty evaluations”. Assessment was least frequently cited to be involved in “participating in institutional assessment efforts”, “research project management”, and “leading departmental assessment efforts”. However, all tasks had a mean rating of more frequent involvement than occasionally (that is, > 2 on a 1-5 Likert scale), and 66% of tasks were reported to involve assessment “sometimes” or more frequently. These findings were supported by qualitative responses to an open-ended question which asked faculty “how is assessment commonly used in higher education?”. Much like the survey data, the open-ended responses most frequently relayed themes of faculty management and decision-making across subthemes involving hiring, tenure, and promotion processes, faculty teaching evaluations, and assessing awards. Faculty are directly impacted by such processes (e.g., in terms of workload, incentives, and career progression), and as such it is reasonable that faculty consider assessment important in these processes.

The earlier scoping review identified that a common thread among early investigations into AL in HE are the specific impacts of local assessment contexts on individual AL, such as how colleagues’ approach to assessment can moderate assessment practice (Medland, 2019). The finding that faculty from a range of diverse programs tend to involve assessment in specific HE tasks supports the argument that existing conceptualizations of AL do not account for all aspects of HE assessment work. This notion has been touched on in previous literature. For example, Arreola, Theall, and Aleamoni (2003) described how HE faculty require the competencies of a base profession as well as the contextually combined elements “from a variety of several other professional arenas” (p. 2). In addition, Wellman (2013) exposed a dearth of literature pertaining to connections between HE assessment costs and outputs. Schoepp and Tezcan-Unal (2017) found that more work was needed for faculty to perceive how their assessment efforts connect to course or program improvement. Furthermore, Medland (2019) explained the need for development of discourse surrounding HE assessment such that appropriate language may be shared and used to discuss AL in HE. The findings from the present study incorporate these disparate incremental deficiencies to explain that, en masse, HE assessment practices occur against a significantly different contextual backdrop than K-12 educational contexts.

These findings emphasized how within these diverse HE contexts the underlying necessary competencies to be ‘assessment literate’ involve different knowledge and practices

than in K-12 contexts. In HE, AL must acknowledge wider assessment utilization than that implied in existing AL conceptualizations. It was found that assessment is frequently used for some tasks shared with K-12 contexts such as assessing student learning within courses, faculty evaluations based on peer assessment, and engaging in feedback based on assessments. This was expected given that assessment theory must have elements of underlying stability that are consistent across contexts (Beebe, 2010). However, assessment was also reported to be frequently used in everyday faculty tasks specific to HE contexts like tenure/promotion/hiring processes, evaluating course material and research products, and using assessment to evaluate academic programs. Since assessment is involved in some tasks that both do and do not overlap with K-12 contexts, the same competencies cannot be implied when using the term AL in both contexts. As emphasized by Norcini et al. (2011), the most effective assessment practices reflect specific purposes in particular contexts. In HE, the assessment knowledge implied in when using the term AL must extend to appropriate awareness, understanding, and implementation of empirically supported assessment techniques specific to HE (e.g., Adachi, Tai, & Dawson, 2018; Beebe, Vonderwell, & Boboc, 2010).

The resistance to incorporating non-student assessment when operationalizing AL may have been an intentional omission within the scholarship of assessment. Ewell and Cumming (2017) explained how avoiding excessive professionalization of HE assessment promoted connections with the scholarship of teaching. Despite this, the present exploratory survey findings demonstrated how assessment is utilized in HE for a range of tasks tangential to student learning assessment such as faculty tenure/promotion decisions, faculty evaluations based on student/peer assessment, or evaluating course materials or research products. For tasks such as these, some knowledge and skills emphasized for student learning assessment are less important or else inappropriately framed. For example, surveyed faculty reported assessment as only sometimes involved in participating in feedback processes. In literature, feedback may be framed in terms of dialogue between assessor (faculty member) and assessed (the student), implying feedback as a process that occurs exclusively in the context of using assessment data to improve student learning. However, in HE feedback is equally important between faculty members and institutional administrators, so that individual faculty members may “see how their assessment reports and data [are being used for course or program improvement]” (Schoepp & Tezcan-Unal, 2017, p. 314). From the tasks in HE in which assessment is used in HE, it is clearly inadequate to use the term “assessment literacy” to describe the necessary skills, knowledge, and contextual competency necessary for efficacious assessment practice in HE contexts.

This finding therefore contributes to AL literature by beginning to operationalize how the term “assessment literacy” can be applied to describe the necessary skills, knowledge, and contextual competency necessary for effective HE assessment. Existing conceptualizations of AL acknowledge the contextually situated nature of AL, but consistently emphasize assessment of student learning as the dominant function of assessment described by AL (e.g. Xu & Brown, 2016). In sum, findings from research question one support the argument that HE faculty involve assessment in a range of HE-specific tasks that have not been accounted for in existing

conceptualizations of AL. Therefore, a specific conceptualization of AL for HE is necessary to encapsulate and describe HE assessment practice in broader literature.

7.2 Research Question Two

How can common higher education assessment tasks be organized and labelled?

With a range of involved professional tasks involved in HE identified, it was found that assessment tasks in HE demonstrated some evidence of being empirically separable into the categories Faculty Service, Feedback/Evaluation, and Program Assessment. The factor structure of the tasks from this survey demonstrated strong evidence of internal consistency including Cronbach's alpha (0.916) and McDonald's omega (0.921), though it was emphasized that the factor analysis was conducted using a small sample with low generalizability. In general, each particular HE task investigated loaded on to these factors in a way that was expected from the theoretical frame established from the scoping review thematic data. Notably, student assessment was not reflected as a distinct category in the factor analysis, but elements of student assessment did group thematically in the qualitative survey items. This finding means that the tasks for which HE faculty involve assessment are not disparate. Instead, the tentative factor structure that emerged may pragmatically describe HE assessment processes, demonstrating that assessment is used in HE with differing frequency for different kinds of tasks. This was also evident from the correlations among the three factors. Program Assessment showed moderate correlation with Faculty Service and Feedback/Evaluation suggesting a spread and overlap of assessment related tasks in HE. This may represent a distinct feature for AL in HE wherein decisions and procedures pertaining to Program Assessment can affect practices related to Faculty Service and Feedback/Evaluation.

Another notable observation was how the majority of assessment related tasks in HE, as reported by the participants, involve some form of Faculty Service (i.e., it accounted for 28.5% of the total variance). The two other factors contributed almost equally to the notion of AL in HE (i.e., Feedback/Evaluation accounted for 16.9% and Program Assessment accounted for 14.9% of the total variance). The observation that faculty seem to perceive assessment-related activities in HE to reach beyond assessing student learning, alongside with the correlations among factors, implies that AL in HE has its own distinct characteristics requiring more detailed attention. These findings conform with previous literature emphasizing the need for treating AL in HE differently than K-12 setting (e.g., Arreola, Theall, & Aleamoni, 2003; Medland, 2019; Schoepp and Tezcan-Unal, 2017; Wellman, 2013). This convention was also supported by qualitative survey data. One respondent astutely emphasized the multiplicity of context-dependent meanings the term *assessment* can have by situating it within their professional frame:

“We are unique. We are a professional college so assessment can mean assessment as a teacher/educator or a professional veterinarian. These are not the same assessment.”

Comments such as these highlight the distinction between aspects of HE AL necessary for student assessment versus other assessment purposes in HE settings. This contention underscores

that AL as it has been defined in K-12 contexts cannot be seamlessly ported to more complex HE contexts.

To answer research question two, common HE assessment tasks can be organized as they relate to Faculty Service, Feedback/Evaluation, and Program Assessment. As well, student assessment can be considered a utility of HE assessment that was assumed to be distinct and was not uniquely elevated in this study. Without affording it special status as the most important kind of assessment, student assessment was not reflected as a distinct category in the factor analysis. Elements of student assessment did group thematically in the qualitative survey items. Despite a prompt to identify uses for assessment outside of assessing learning, management of student learning was identified in relation to discipline-specific student assessment opportunities including experiential learning, clinical settings, and authentic assessments. Critically, this interpretation is made with awareness that further categories of HE assessment tasks represented by facets of HE assessment not queried by this survey likely exist. Regardless, these findings provide valuable information about how assessment is used in HE and as such hint at what competencies may be necessary to considered assessment-literate in HE contexts. This observation thus has exciting value as a jumping-off point from which to further understand how the everyday work of HE faculty can be acknowledged and categorized.

7.3 Research Question Three

Research question three investigated traditional statistical relationships between HE assessment tasks and theorized components of AL for HE. Pearson correlations showed significant relationships between different arrangements of assessment task factors and theorized components of AL for HE.

Faculty Service correlated positively with conceptions, assessor identity, and decision-making. This means faculty who more frequently involved assessment in Faculty Service tasks were also likely to conceive of assessment as positive or useful, identify themselves as assessors, and to consider assessment necessary in decision-making processes. This finding means that assessment involvement in Faculty Service is related to important aspects of AL in HE as identified by the scoping review. This indicates that aspects of everyday HE work not considered by previous conceptualizations of AL are nonetheless related to AL. For example, Xu and Brown's (2016) TALiP model places identity as assessor as the highest-order indicator of teacher AL in practice. The present survey findings demonstrated that those who involve assessment in Faculty Service also tend to identify themselves as assessors. Although not causative, the finding that assessment involvement in Faculty Service increases with expected indicators of AL for HE is important for the argument towards a conceptualization of AL unique to HE contexts.

The factor Feedback/Evaluation was positively correlated with knowledge, metacognition, and decision-making. This means faculty who more frequently involved assessment with Feedback/Evaluation also tend to consider themselves to have sufficient theoretical and applied assessment knowledge. They tended to agree that they reflected on their own assessment knowledge and practice and considered assessment a major step in decision-

making processes. Assessment involvement in Feedback/Evaluation thus appeared to rise alongside consideration of technical and theoretical aspects of enacting assessment processes correctly or completely. Feedback/Evaluation tasks seem to relate to opinions about ‘closing the loop’ by using data from an assessment to monitor or make judgements about effectiveness (e.g., Banta & Blaich, 2011). It thus makes sense that those involving assessment with feedback would be more engaged in metacognitive self-regulation regarding reflection on their own assessment knowledge. Those using assessment for these purposes are considerate of elements of AL related to sufficient knowledge. They consider themselves to possess sufficient and appropriate theoretical knowledge and skills to efficaciously perform related tasks. As well, such faculty engage in metacognitive reflection on such sufficient knowledge and skills. They consider the knowledge and skills in assessment and self-monitor if they are being used in appropriate capacities. Last, they tend to involve assessment data in decision-making, and in so doing reinforce the necessity of assessment processes by ensuring assessment data is mobilized towards improvement or accountability purposes.

Program Assessment correlated positively with knowledge, metacognition, and assessor identity. This means faculty who are involved with Program Assessment considered theoretical and applied knowledge and skills important, tend to reflect on their own assessment practice, and consider themselves to be assessors. This observation also indicates that specific components of AL are implicated with particular kinds of assessment tasks. Further, it could be the case that these general HE tasks invoke components of AL in different ways than other assessment contexts (i.e., K-12). This provides support for developing understanding of AL for HE along a pathway that favours contextual profiles involving different amounts of particular components of al, akin to the approach utilized by Kremmel and Harding (2020) in their model for LAL. Were data collected to this end, it would be possible to compare how AL is constituted for particular task profiles: for example, Faculty Service could be compared to Feedback/Evaluation to compare which AL components are considered more important or utilized more frequently.

Along this line of thought, it is useful to observe the overlaps and differences among how AL components are correlated among the factorial structure identified here. Faculty Service and Program Assessment both significantly correlated with assessor identity but differed on other correlations. Both of these factors correlated with components shared in Feedback/Evaluation, which was correlated with knowledge, metacognition, and decision-making. Here, it was seen that each type of task involves a profile of components that may be unique to that particular type of task, or else shared with some tasks but not others. This demonstrates the necessity and utility of conceptualizing AL for HE: different types of tasks rely on unique arrangements of AL components emulating Kremmel and Harding’s (2020) developmental approach. This further indicates that AL is manifest contextually as espoused by Willis, Adie, and Klenowski (2013) and supported by more recent AL developments (DeLuca et al., 2018; Looney et al., 2017; Medland, 2019).

However, it is important to highlight how this study examined suspected AL components as equal and separate, when in fact it is likely that some components underlie others. Assessment

knowledge underlies all other AL components in Xu and Brown's (2016) TALiP model, and identity-as-assessor is considered the top of a hierarchy that emerges from the interaction of all the components nested beneath it. In Chapter Four, AL themes emergent from the scoping review were compared to TALiP in detail, finding that some themes of AL for HE align with components of TALiP, such as knowledge or development. Other themes that seem important for AL in HE, such as assessment fluency (Medland, 2019), did not fit neatly within TALiP and indicated the necessity of further investigating elements and features of AL unique to HE contexts. When considering how to relate components of AL for HE in a conceptual model, the results indicate it may be useful to adopt a profile approach like Kremmel and Harding's (2020) LAL rather than a hierarchy like Xu and Brown's (2016) TALiP. Components of AL arrange differently for different tasks, and it may be limiting to attempt to describe AL in terms of skills and knowledge that may not be important for particular tasks. TALiP is a useful conceptualization of AL for a particular context (K-12 teachers) and particular tasks (student learning assessment), but such a structure may be too rigid to accommodate the wide range of assessment uses and contexts in HE.

Overall, these findings are important because they demonstrated how expected elements of AL for HE are indicated in different arrangements for different types of tasks. The correlations showed that unique factors of HE professional tasks were positively associated with different constellations of theorized components of AL for HE. Thus, HE faculty involve assessment differently in different types of tasks, and tend to agree with statements about assessment in different ways depending on the types of tasks they are involved in. As such, it seems plausible that elements of faculty AL may be invoked differently for assessment-related tasks in HE not directly supporting student assessment, countering existing conceptualizations of AL formulated in K-12 settings that focus on student assessment tasks.

7.4 Research Question Four

Research question four investigated how HE assessment and its involvement in HE professional work was uniquely construed by HE faculty. Pearson correlations among theorized components of AL for HE suggested that faculty who know more about HE assessment also tend to value and reflect about assessment. Those who conceive of assessment as useful also tend to consider themselves as experienced, engaged assessors and consider assessment important to make decisions. Those who consider themselves experienced assessors also tend to feel confident understanding and using assessment language and grammar, and tend to participate in opportunities to develop their AL.

These correlations can be applied to investigate if an intuitive practical conceptualization of AL may be implicated by this survey data. Although not all components correlated with each other, the following image visualizes how components of AL for HE correlated significantly within the scope of this limited sample:

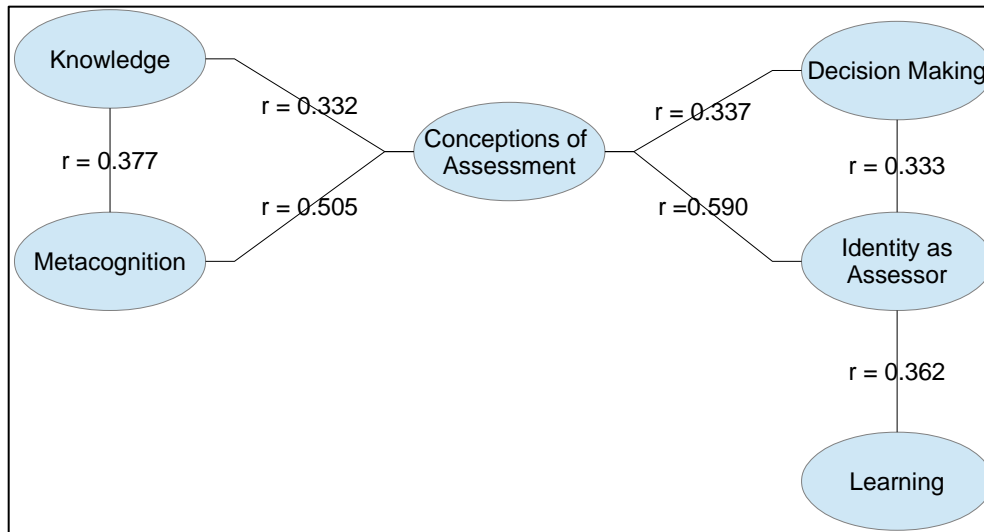


Figure 7.1 Correlations among theorized components of AL for HE.

Knowledge, metacognition, and conceptions are visualized as correlated in a cluster together. Another cluster of correlations appears among conceptions, decision-making, and identity-as-assessor. The first clusters of components seems to represent cerebral AL skills like theoretical knowledge, self-reflection, and opinion which could be considered intrapersonal indicators of AL. The second cluster of components seems to represent interpersonal indicators of AL that are more conspicuous. Conceptions seem to bridge these different clusters. Outside of these, learning correlated only with identity as assessor; fluency did not significantly correlate with other components. This finding means that expected components of AL for HE are related to one another, but data from this survey were not sufficient to investigate the direction and nature of these relationships. While some patterns appear to emerge from the correlations, more evidence about the directionality of this data is necessary to make meaningful arguments about a conceptual structure of AL for HE. Instead, this limited survey data is useful to begin an investigation of the how assessment involvement in particular HE tasks is related to particular statements about assessment.

To this end, independent t-tests using True/False items as grouping variables demonstrated how faculty involved in different types of assessment tasks responded in different ways to different statements. These items were scaled dichotomously for with the intention of grouping respondents based on how they perceived tasks as involved in assessment. Dichotomous items only have minimal variability, and these items only have two levels of covariation: true or false (DeVellis, 2017). This limits scale variance for these items, but for the purpose of this analysis it was useful to sort respondents based on their agreement to opinion statements. This process bore several insights. Faculty who felt capable of applying their assessment knowledge and considered it necessary for making decisions were more likely to involve assessment more frequently in all types of HE assessment tasks. Similarly, faculty who discerned opportunities to apply assessment knowledge involved assessment more frequently in more routinized assessment tasks, but not Faculty Service tasks. Additionally, those who felt

their assessment effort yielded worthwhile results tended to involve assessment more frequently in tasks typical of HE professional work. Taken with the previous result, it may be possible that faculty who perceive their efforts as worthwhile typically involve assessment more frequently in Faculty Service tasks and perceive opportunities to involve assessment in other kinds of tasks.

Finally, as faculty gained experience using assessment they seem more willing to involve it in a greater variety of HE tasks outside of Feedback/Evaluation. It may be that more experienced assessors have attained higher ranks within their institution and have a more diverse profile of professional responsibilities. In a similar way, confident faculty may perceive the assessment involved in feedback and evaluation tasks to a greater degree than non-confident faculty.

These findings elucidate the contextual complexity of AL for HE. Faculty with different assessment-related opinions involved assessment differently in various kinds of HE tasks. Medland (2019) pointed out that contemporary AL literature tends to apply deficit models in which AL is investigated as either present or absent in different amounts. The present data illustrate how such deficit models are inadequate to account for nuances in the AL of HE faculty. Merely describing faculty as ‘assessment literate’ or ‘not assessment literate’ overlooks the established contextual nuances of how faculty understand and use AL. This is important as it underscores the importance of a model for AL in HE that accounts for how faculty already understand and use assessment in context, but also provides targeted pathways for further development.

7.5 Discussion Summary

In the early stages of this study, assessment was identified as an onerous, time-consuming, and yet critical part of everyday academic life (Huber & Skedsmo, 2016). A working definition was established describing HE assessment as the process of gathering, documenting, analyzing, and interpreting empirical evidence to make decisions related to situations like (but not limited to) assessing student learning, hiring/advancement, program evaluation, admission, and promotion. From this definition, this thesis set out to establish how HE assessment is invoked in different kinds of HE work and if a generalized conceptualization of AL for HE could be established to describe and compare a range of diverse HE contexts. Following a scoping review conducted for this purpose, it was speculated that AL in HE may have an element of selective professional knowledge where assessment knowledge and skills are utilized for necessary tasks for particular purposes. The present findings from analyzing survey data demonstrate that assessment is implicated in a range of HE tasks that are empirically separable and which have not been considered in previous literature regarding AL. Further, these different HE task factors are related with varying strength to different arrangements of components of AL. As well, components of AL that were speculated from thematic analysis of a body of literature related to AL in HE were demonstrably related to one another in various ways. Together, these findings indicate that AL may be explained by examining composite sub-concepts in relation to one another, and further that such a model for AL is related to different tasks in different ways.

This underscores the idea that particular purposes for assessment in HE may be nested within a more general conceptualization of AL.

By conceptualizing AL for HE from the ground up by examining what AL components are common *across contexts*, a framework for comparison and development may then be applied to specific types of tasks or purposes in which assessment is used. This echoes approaches seen in contemporary conceptualizations of LAL wherein constituent components are present in varying amounts for different roles related to language assessment (Kremmel & Harding, 2020). AL has not been conceptualized this way for HE, but the present results suggest that emulating such a model for AL for HE shows deep promise. Developing an AL framework that generally describes aspects of HE assessment but also specifies how assessment is utilized for specific professional purposes in HE would be especially useful in terms of creating opportunities for faculty development that is targeted and useful in a wide range of specific contexts. For example, a range of operationalized components of AL applicable to a variety of tasks within a variety of HE contexts would be useful to tailor AL professional development towards knowledge and skills more valuable within that particular context.

Data gathered during this thesis were not sufficient to establish a sound model for comparison in this way. Rather, in composite this study can be considered evidence that such a conceptualization shows promise for greater future understanding of AL in HE, what may constitute it, and directions for how such a model could be utilized.

7.6 Limitations

Findings from the present study should be interpreted with acknowledgement of some significant limitations throughout the research process as described here.

The entire thesis faced limitations as it used exploratory modes of research. Although the inherent lack of specificity of the construct demanded exploratory methods, investigation of the research questions led to an expanse of further questions and directions for investigation. The scope and nature of the project evolved as the investigation progressed in response to the emergent data, and at times the process was akin to navigating a labyrinth with a blindfold. Some promising directions ended up redundant, whereas processes assumed to be methodological and straightforward became the biggest problems of the project (i.e., gathering a representative sample). As well, the sheer breadth of scope with which educational research as a task has approached assessment unearthed a plethora of possible approaches to investigating the construct. Though the thesis adopted one method that aligned with the pragmatic epistemology of the project, it is possible that any number of other approaches may have been more fruitful in specifying the construct. This is an inherent convention to adopting pragmatic exploratory methods. It is hoped that the challenges and successes described throughout the thesis may be useful guiding points to future investigations into the construct of AL for HE.

The scoping review was a challenge in of itself given the task to gather literature on an unspecified construct. The PRISMA scoping review methods used rely on using clear, specific search terms which was difficult as the spirit of the thesis was to bring clarity to such terms. For

the review, operationalization of the terms was inherently unclear based on the broad ways in which the term *assessment* is used in different contexts and sub-contexts. This is even more true for the term AL; this limitation was, in a sense, the initial research question for this study. The construct of “Assessment Literacy in Higher Education” is not well defined or understood, and this project initially set out to determine how AL may be conceptualized differently for this particular context. As such, a broad approach to inclusion was necessary for the scoping review, but this yielded an ever-expansive body of assessment-related literature, most of which did not relate to the project at hand. Even after data reduction, deep thematic analysis of each individual paper was not feasible due to time and resource constraints. Due to the scope of the data that were yielded, it is possible that aspects of assessment that may be important in AL for HE were not adequately sampled by the scoping review, and as such were overlooked throughout the rest of the project.

AL has been primarily conceptualized in relation to school-based settings, but such settings were intentionally excluded from the scoping review to focus on how the concept is used in HE. As it has yet to be determinately conceptualized for HE settings, the projected definition of AL may not have accurately represented the construct at the outset of the project. Though a broad and lenient set of search terms were used, it is possible that pertinent aspects of AL for HE were not captured by the search terms. Conversely, it is possible that the wide net cast by a scoping review incorporated terms unrelated to AL for HE settings. Furthermore, it is possible that elements, features, and aspects of AL for HE settings were overlooked or masked by other elements within the literature. This risk is reasonable given the purposes of a scoping review to synthesize information from a broad search of literature into a comprehensible representation of the most recognizable and characteristic features.

In similar ways, the survey study was limited by the non-specificity of the construct under investigation. *Assessment* is a broad and highly general term with many interdisciplinary applications, and as such the developed questionnaire was biased to reflect a definition of assessment most directly relevant to the research at hand. Despite this, *assessment* is defined and in a myriad of nuanced ways; this became increasingly evident during the expert review of the questionnaire development wherein various assessment experts rated elements of HE assessment in polarizing ways, thus eliminating them from the questionnaire due to conceptual incongruence. Additionally, the survey was disseminated to faculty from a range of research disciplines where the same issue with definitional incongruence applies. For example, a large contingent of respondents had backgrounds in Dentistry, where the term *assessment* is utilized and understood differently to those with backgrounds in Education or Psychology. This risk was mediated by providing a functional definition of higher educational *assessment* at multiple points throughout the questionnaire. Ironically, any interpretation of these data must account for possible discrepancies in participant *assessment literacy* based on their own professional experience and disciplinary context.

Further limitations are acknowledged in the process of survey development. The useful and concise guidelines promoted by DeVellis (2017) informed the process of developing the

scale, but further guidance specific to the project was necessary. Given that Kremmel and Harding (2020) achieved a survey with practical application to those involved in LAL work, the survey development process they described was considered in development of the AL for HE survey. This created some confusion in terms of survey purpose, which was highlighted by comments by five expert reviewers. A panel of more reviewers may have impacted how items were deemed useful and would have provided even better qualitative feedback regarding the content universe. Regardless, review of the content domain of survey items suggested overlap regarding item phrasing and it seems likely that they do not meaningfully differentiate the constructs they intend to represent. Also, items exploring survey themes were scaled dichotomously. If these items had been scaled in a way to explore their factor structure (e.g., Likert-style) as opposed to being used as scored components and grouping variables, they may have represented greater variation in how the sampled faculty perceive thematic components of AL for HE.

The survey study was limited in ways that are typical for research methods incorporating online surveys. It is likely that gathered data did not accurately represent the views of all faculty members within the population. Faculty could elect to respond based on any number of factors including conceptions of assessment, filling time, or happening to receive the recruitment email at a convenient moment. Further to this, faculty could discontinue the survey at any given point. The unclear content universe created some analytic limitations for the thesis. Given the primacy of student assessment in existing literature regarding AL itself, it seems likely that responses to a survey with more items reflecting specific HE student assessment processes may cluster to form an additional factor regarding student learning assessment.

One of the most pressing limitations was the nature of the sample; HE faculty proved a difficult population to elicit response from. Important decisions such as content reduction, item removal, and number of factors to extract were made on the basis of a small convenience sample. As described throughout analysis, the sample size fell significantly short of the target n of 200 respondents. Recommended sample sizes for factor analysis begin liberally at $n = 150$, with larger samples and a large item-to-participant ratio encouraged. The present sample resulted in an only 1:3 item-to-participant ratio which is far short of larger ratios recommended to increase statistical power (Bujang et al., 2012). The inability to recruit a sufficient sample impacted the generalizability of the findings and it is possible the statistical analyses misrepresented the construct due to the limited opportunities for variation to be represented (DeVellis, 2017). If a larger sample had been achieved, it is likely that the underlying response patterns may have arranged differently. A different number of factors may have been more optimal, and responses to items may have correlated differently. Because a larger sample would be more representative, it is imperative that the findings presented here are interpreted cautiously and not used as a basis for future work. Rather, the findings should be treated as exploratory only. Interpreting the findings with consideration for the small sample was still fruitful, especially in terms to the qualitative data yielded. As well, the sample adequacy was deemed appropriate based on measures such as the KMO index and Bartlett's test of sphericity which are recommended when

proceeding with an EFA using a sample with a low item-to-participant ratio (William, Onsman, & Brown, 2010). For these reasons, the EFA proceeded and yielded useful, if not generalizable, insights that confirm the value of future inquiry into AL in HE.

The survey sample had also been gathered for its convenience by recruiting participants from the local institutional context which required only the ethical approval from the local ethical review board. Sampling from a wider swath of universities would have required ethical approvals from each given institution which was not feasible within the scope of this graduate thesis project. Further, there was much contextual variation within the convenience sample: respondents ranged across a wide variety of programs and departments, and as such there was much room for disagreement about assessment use among faculty from different programs. Because of the small sample size, responses were not interpreted with regards for program of origin. This was a major shortcoming as the project sought to understand AL in context but could not make fair interpretations regarding disciplinary context. A larger sample may have mitigated this, but even a sample of at least 200 may not have achieved adequate representation across the variety of programs represented. Another approach to mitigate this may have been to draw a sample from only a few academic departments. This approach had been used by Davies and Taras (2018) who compared the assessment knowledge of and skills of HE faculty in only two departments. This may have allowed for a more consistent comparison and is a direction to be considered for future inquiries.

One further limitation of this study was the chronological and historical context under which the data were gathered. The study was conducted during the COVID-19 global pandemic, which prefaced sudden and dramatic shifts to how HE faculty perform professorial and academic roles. In Canada, HE shifted to a primarily online delivery format and as such the outcomes of this study are impacted by the atypical functioning of the HE sector. The target population was required to accommodate to rapid shifts in how they were expected to perform their everyday work. In particular, the survey development and delivery phases were impacted by COVID-19 as it delayed the timeline of the project. The survey was not piloted as thoroughly as possible due to time constraints on the project. Nevertheless, other considerations for survey quality, such as expert review of survey content, were pursued to assure the psychometric quality of the survey instrument.

7.7 Future Directions

Because AL may be indicated in different ways for different tasks, it will be necessary to attempt to understand how components of AL relate to one another within a conceptual framework. As different kinds of tasks in different situations draw on different aspects of AL, it may seem fruitless to strive for a conceptualization of AL for HE that can be used to explain a phenomenon with such contextual fluidity. This thesis contends it is critical to develop a pragmatic conceptualization from the ground up that emerges from patterns of assessment skills and practices necessary in diverse HE contexts (Rorty, 2009). In this way, general features of AL

in HE that describe *most* AL contexts and tasks can be applied to particular uses and differences/similarities can be compared and evaluated.

Some future directions have been alluded to throughout the discussion and the limitations sections. Pressingly, the survey and analysis may have been more effective overall with a more representative sample and further opportunity to collaborate with other experts throughout the process. Likely, a multidisciplinary team using assessment for a range of purposes may be able to achieve a survey that explores AL in HE in a more contextually situated manner. Further, a different sampling strategy ought to be considered in order to achieve a sample with stronger psychometric properties. As well, future studies ought to consider operationalizing assessor context more rigidly in terms of demographic variables such as home discipline or faculty ranking. Probability sampling methods may be utilized to this end. This may yield a sample from which stronger generalizations can be drawn and from which more useful recommendations can be made. At the very least, this study can be considered evidence that such a conceptualization shows promise for greater future understanding of AL in HE, what may constitute it, and directions for how such a model could be utilized.

Many future directions can be considered on the basis of the work presented in this thesis. From the outset, there had been five guiding objectives for this thesis. First, to develop a comprehensive understanding of HE assessment by determining encompassing elements, features, procedures, and actions that have been implicated in academic research. Second, to establish a thematic framework synthesizing research literature on AL in HE. Third, to establish preliminary empirical support for AL in HE as conceptualized by research literature. Fourth, to develop guidelines supporting HE administrators and academic developers in design and delivery of professional development programs in AL. Fifth, to disseminate research findings and mobilize knowledge of AL in HE to education community contexts. With the size of the project, the latter two of these objectives have not yet been achieved and stand as directions for imminent future work on this topic.

The first and second of these purposes was achieved using a scoping review. The scoping review yielded a useful collection of academic sources related to HE assessment and AL. Further work could be done to synthesize data from this library to approach the conceptualization of AL for HE from different theoretical standpoints than those presented here. It is highly likely that a more extensive thematic analysis or a systematic review may yield more specific insights into how AL for HE has been discussed directly or tangentially in academic literature to date. In addition, the scoping review usefully revealed a gap in understanding in how HE faculty involve assessment in their everyday work outside of student learning assessment. A thorough academic account of how HE faculty use assessment in their everyday work was considered imperative for developing a contextually situated framework for understanding AL in HE, but the present review approached these tasks in a manner of convenience by thematically coding for such tasks from within the scoping review library. Further scoping reviews or systematic reviews may be targeted with this question in mind from the outset.

The third objective had been to establish preliminary empirical support for AL in HE as conceptualized by research literature. As discussed, the scope of this objective shifted following the scoping review. Prior to establishing a thorough conceptualization of AL for HE, it was deemed pertinent to further understand the context of HE assessment with consideration to how faculty use assessment. The survey project had initially been constructed to attempt both of these tasks; following the initial survey development and feedback from reviewers it was decided to focus the investigation on how HE faculty involved assessment in their everyday work. This was then correlated with their agreement or disagreement with statements reflecting themes of AL for HE as elicited by the scoping review. Usefully, the qualitative survey data provided valuable thematic insight as to how assessment is specifically invoked in HE contexts. For this sample, assessment was seen as heavily involved in faculty decision-making, student decision-making unrelated to learning, program assessment, resource management, and faculty service, as well as some other tasks. This qualitative data is extremely valuable to guide further explorations into how HE faculty use assessment, and what they need to know to be effective at it.

In this way, the objective to establish empirical evidence of a framework of AL for HE was not yet achieved, but important steps were taken towards enhancing understanding of that task itself. With this in mind, future investigations can focus on elucidating how assessment is differently involved in HE work as well as seeking to understand how different functions of HE assessment may be associated with one another. Only with this understanding can a representative and subject-oriented conceptualization of AL for HE be achieved.

References

- Academy of Process Educators. (2016). Assessment vs. evaluation. *International Journal of Process Education*, 8(1), 53-58. Retrieved from <http://www.ijpe.online/2016/color033116sm.pdf>
- Adachi, C., Tai, J. H.-M., & Dawson, P. (2018). Academics' perceptions of the benefits and challenges of self and peer assessment in higher education. *Assessment & Evaluation in Higher Education*, 43(2), 294–306. <https://doi.org/10.1080/02602938.2017.1339775>
- Alsobrook, M. (2010). *Assessment of student learning outcomes in a university setting: The case of the university of Texas at Dallas* (Order No. 3421558). Available from Education Database; ProQuest Dissertations & Theses Global. (753901107). Retrieved from <http://cyber.usask.ca/login?url=https://search.proquest.com/docview/753901107?accountid=14739>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association. Retrieved from <https://www.testingstandards.net/open-access-files.html>
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. Lincoln, NE: Buros Center for Testing.
- Apple, D. K., Krumsieg, K. (1997). *Process Education Teaching Institute: Handbook*. Pacific Crest Software. <https://books.google.ca/books?id=bZmatwAACAAJ>
- Arreola, R. A., Theall, M., & Aleamoni, L. M. (2003). *Beyond scholarship: Recognizing the multiple roles of the professoriate*. Paper presented at the 83rd Annual Meeting of the American Educational Research Association. Chicago: April 22. Available at: www.cedanet.com/meta/Beyond%20Scholarship.pdf
- Assessment. (2020). In *Lexico Online Dictionary*. Retrieved from <https://www.lexico.com/definition/assessment>
- Assessment for Learning in Canada. (2018, May 19). Retrieved September 22, 2020, from <https://cafln.ca/resources/assessment-for-learning-in-canada/>
- Baker, B. (2016). Language assessment literacy as professional competence: The case of Canadian admissions decision makers. *Canadian Journal of Applied Linguistics*, 19(1), 63–83. <https://journals.lib.unb.ca/index.php/CJAL/article/view/23033>
- Baker, B. A., Tsushima, R., & Wang, S. (2014). Investigating language assessment literacy: Collaboration between assessment specialists and Canadian university admissions officers. *Language Learning in Higher Education*, 4(1), 137–157. Education Database. <https://doi.org/10.1515/cercles-2014-0009>
- Bandy, J., Bartel, A. S., Clayton, P. H., Gale, S., Mack, H., Price, M., Nigro, G., & Stanlick, S. (2016). Values-engaged assessment: Reimagining assessment through the lens of democratic engagement. *Michigan Journal of Community Service Learning*, 23(1), 96–101.

- Banta, T. W., & Blaich, C. (2010). Closing the Assessment Loop. *Change: The Magazine of Higher Learning*, 43(1), 22–27. <https://doi.org/10.1080/00091383.2011.538642>
- Bastanfar, A. (2009). Alternatives in assessment. *Karen's Linguistic Issues*. Retrieved from https://dphu.org/uploads/attachements/books/books_4732_0.pdf
- Beckwitt, E. G., Silverstone, S., & Bean, D. (2010). Creating a culture of academic assessment and excellence via shared governance. *Contemporary Issues in Education Research*, 3(2), 35-48. Retrieved from <http://cyber.usask.ca/login?url=https://search.proquest.com/docview/196350539?accountid=14739>
- Beebe, R., Vonderwell, S., & Boboc, M. (2010). Emerging patterns in transferring assessment practices from F2f to online environments: EJEL. *Electronic Journal of E-Learning*, 8(1), 1-12. Retrieved from <http://cyber.usask.ca/login?url=https://search.proquest.com/docview/856420216?accountid=14739>
- Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., Smith, K., Timperley, H., Volante, L., & Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education*, 13(1), 117–140. <https://doi.org/10.1177/1478210314566733>
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Blaich, C., & Wise, K. (2011). *From gathering to using assessment results: Lessons from the Wabash national study*. National institute for Learning Outcomes Assessment. <https://www.learningoutcomesassessment.org/wp-content/uploads/2019/02/OccasionalPaper8.pdf>
- Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill.
- Bloxham, S., den-Outer, B., Hudson, J., & Price, M. (2016). Let's stop the pretence of consistent marking: Exploring the multiple limitations of assessment criteria. *Assessment & Evaluation in Higher Education*, 41(3), 466–481. <https://doi.org/10.1080/02602938.2015.1024607>
- Blumenstein, M. (2015). The generic-discipline dilemma: Learning advising in quantitative literacy. *International Journal for Researcher Development*, 6(1), 24–39. <https://doi.org/10.1108/IJRD-09-2014-0032>
- Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research In Psychology*, 3(2), 77-101.
- Brumfit, C. (2010). Literacy or literacies? Academic identities in a language-sharing world. In Blue, G. (Ed.), *Developing Academic Literacy*. Bern, Switzerland: Peter Lang.
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, 32, 513–531. <https://doi.org/10.1037/0003-066X.32.7.513>

- Brookhart, S. M. (2011). Educational Assessment Knowledge and Skills for Teachers. *Educational Measurement: Issues and Practice*, 30(1), 3–12.
<https://doi.org/10.1111/j.1745-3992.2010.00195.x>
- Brookhart, S., Guskey, T., Bowers, A., McMillan, J., Smith, J., Smith, L., Stevens, M., & Welsh, M. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803–848.
<https://doi.org/10.3102/0034654316672069>
- Brown, G. (2004). Teachers' Conceptions of Assessment: Implications for Policy and Professional Development. *Assessment in Education Principles Policy and Practice*, 11, 301–318. <https://doi.org/10.1080/0969594042000304609>
- Brown, G. T. L., Hui, S. K. F., Yu, F. W. M., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research*, 50(5), 307–320.
<https://doi.org/10.1016/j.ijer.2011.10.003>
- Brown, G. T. L., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. *The Spanish Journal of Psychology*, 15(1), 75–89.
https://doi.org/10.5209/rev_SJOP.2012.v15.n1.37286
- Bujang, M. A., Ghani, P. A., Soelar, S. A., & Zulkifli, N. A. (2012). Sample size guideline for exploratory factor analysis when using small sample: Taking into considerations of different measurement scales. *2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE)*, 1–5. <https://doi.org/10.1109/ICSSBE.2012.6396605>
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002, October). Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers. In *Annual meeting of the mid-western educational research association, Columbus, OH*.
- Cattell, R.B. (1978) *The Scientific Use of Factor Analysis in Behavioral and Life Sciences*. Plenum, New York.
- Caudle, L. S. (2014). *Exploring Community College Faculty Involvement in Assessment of Student Learning* (1559962130) [Ed.D., University of Arkansas].
<http://cyber.usask.ca/login?url=https://search.proquest.com/docview/1559962130?accountid=14739>
- Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 83(1), 1–6. <https://doi.org/10.1080/00098650903267784>
- Clark, R., & Filinson, R. (2011). Kicking and screaming: How one truculent sociology department made peace with mandatory assessment. *Teaching Sociology*, 39(2), 127–137.
<https://doi.org/10.1177/0092055X11400439>
- Cohen, J. (2004). Assessment ... yours, mine, and ours. *Journalism & Mass Communication Educator*, 59(1), 3–6. Education Database.
<http://cyber.usask.ca/login?url=https://search.proquest.com/docview/215288566?accountid=14739>

- Cole, A., & De Maio, J. (2009). What we learned about our assessment program that has nothing to do with student learning outcomes. *Journal of Political Science Education*, 5(4), 294–314.
- Collins, J., Richmond, J., & Buzila, H. (2018). *Health assessment: CLPNA self-study course*. College of Licensed Practical Nurses of Alberta: Edmonton.
- Coombs, A., DeLuca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education*, 71, 134–144. <https://doi.org/10.1016/j.tate.2017.12.010>
- Creason, P. J. (2015). *The assessment of student learning outcomes at a California community college: Insight from faculty in a single department* [Ed.D., California State University, Long Beach].
<http://search.proquest.com/education/docview/1697333527/abstract/5F862901901C495CPQ/11>
- Davies, M. S., & Taras, M. (2018). Coherence and disparity in assessment literacies among higher education staff. *London Review of Education*, 16(3), 474–490.
<https://doi.org/10.18546/LRE.16.3.09>
- Deeley, S. J., & Bovill, C. (2017). Staff student partnership in assessment: Enhancing assessment literacy through democratic practices. *Assessment and Evaluation in Higher Education*; Abingdon, 42(3), 463–477. <http://dx.doi.org/10.1080/02602938.2015.1126551>
- DeLuca, C., Coombs, A., MacGregor, S. and Rasooli, A. (2019). Toward a differential and situated view of assessment literacy: Studying teachers' responses to classroom assessment scenarios. *Frontiers in Education* 4, 94. doi: 10.3389/feduc.2019.00094
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, 21(4), 248–266. <https://doi.org/10.1080/10627197.2016.1236677>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2015). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation, and Accountability*, 28(3), 251–272. <https://doi.org/10.1007/s11092-015-9233-6>
- DeLuca, C., Valiquette, A., Coombs, A. J., LaPointe-McEwan, D., & Luhanga, U. (2018). Teachers' approaches to classroom assessment: a large-scale survey. *Assessment in Education: Principles, Policy, & Practice*, 25, 355–375. doi: 10.1080/0969594X.2016.1244514
- DeLuca, C., Willis, J., Cowie, B., Harrison, C., Coombs, A., Gibson, A., & Trask, S. (2019). Policies, Programs, and Practices: Exploring the Complex Dynamics of Assessment Education in Teacher Education Across Four Countries. *Frontiers in Education*, 4.
<https://www.frontiersin.org/article/10.3389/feduc.2019.00132>
- Deygers, B., & Malone, M. E. (2019). Language assessment literacy in university admission policies, or the dialogue that isn't. *Language Testing*; London, 36(3), 347–368.
<http://dx.doi.org/10.1177/0265532219826390>

- DiBiase-Lubrano, M. J. (2018). The importance of assessment literacy for language faculty. *Die Unterrichtspraxis; Cherry Hill*, 51(2), 121-131.
<http://search.proquest.com/education/docview/2151123300/abstract/7E2BC6E0A3045BBPQ/1>
- Dietrich, C. S. (2011). *A Phenomenological Study of Social Science Instructors' Assessment Practices for Online Learning* (889092106) [Ph.D., Walden University]. Education Database.
<http://cyber.usask.ca/login?url=https://search.proquest.com/docview/889092106?accountid=14739>
- DiLoreto, M. A. (2013). *Multi-group invariance of the conceptions of assessment scale among university faculty and students* [Ph.D., The University of Southern Mississippi].
<http://search.proquest.com/docview/1476943655/abstract/5374A7A230EB4EA7PQ/1>
- Dombrowski, S. C. (2015). Purpose of psychoeducational assessment and report writing. In Dueben, R. M. (2015). *Faculty attitudes toward assessment* [Ph.D., Washington State University].
<http://search.proquest.com/education/docview/1776157924/abstract/5F862901901C495CPQ/10>
- Earl, L. M., Katz, M. S., Manitoba, Manitoba Education, C. and Y., & Western and Northern Canadian Protocol for Collaboration in Education. (2006). *Rethinking classroom assessment with purpose in mind: Assessment for learning, assessment as learning, assessment of learning*. Manitoba Education, Citizenship & Youth.
http://www.edu.gov.mb.ca/ks4/assess/wncp/rethinking_assess_mb.pdf
- Earl, L., Volante, D. L., & DeLuca, D. C. (2015). *Assessment for Learning across Canada*. EdCan Network. Retrieved September 22, 2020, from
<https://www.edcan.ca/articles/assessment-for-learning-across-canada/>
- Ebersole, T. (2009). Postsecondary assessment: Faculty attitudes and levels of engagement. *Assessment Update*, 21(2), 1-2, 13-14. <https://doi-org.cyber.usask.ca/10.1002/au.212>
- Egger, M., & Smith, G. D. (1998). Bias in location and selection of studies. *BMJ: British Medical Journal*, 316(7124), 61-66.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2665334/>
- Emil, S. (2011). *Assessment for Improvement in Higher Education: Faculty Perception of and Participation in Program Assessment* [Ed.D., Portland State University].
<http://search.proquest.com/education/docview/889929656/abstract/5F862901901C495CPQ/21>
- Emil, S., & Cress, C. (2014). Faculty perspectives on programme curricular assessment: Individual and institutional characteristics that influence participation engagement. *Assessment & Evaluation in Higher Education*, 39(5), 531-552.
- Evaluation (2020). In *Online Etymology Dictionary*. Retrieved from
<https://www.etymonline.com/word/evaluation>

- Ewell, P. T. (2008). Assessment and accountability in America today: Background and context. *New Directions for Institutional Research*, 2008(S1), 7–17. <https://doi.org/10.1002/ir.258>
- Ewell, P. T. (2009). *Assessment, accountability, and improvement: Revisiting the tension*. National Institute for Learning Outcomes Assessment. <https://www.learningoutcomesassessment.org/wp-content/uploads/2019/02/OccasionalPaper1.pdf>
- Ewell, P.T., & Cumming, T. (2017). Chapter 1. Introduction: History and conceptual basis of assessment in higher education. In Cumming, T. & Miller, D. (Eds.), *Enhancing assessment in higher education: Putting psychometrics to work*. Sterling, VA: Stylus.
- Ewell, P., Paulson, K., & Kinzie, J. (2011). *Down and In: Assessment Practices at the Program Level*. National Institute for Learning Outcomes Assessment. https://www.learningoutcomesassessment.org/wp-content/uploads/2019/02/2011SurveyReport_DownIn.pdf
- Feuerstein, A. (2015). Rituals of verification: Department chairs and the dominant discourse of assessment in higher education. *Journal of Higher Education Theory and Practice*, 15(6), 38–51. <http://search.proquest.com/education/docview/1749280480/abstract/5F862901901C495CPQ/9>
- Fletcher, R. B., Meyer, L. H., Anderson, H., Johnston, P., & Rees, M. (2012). Faculty and Students Conceptions of Assessment in Higher Education. *Higher Education*, 64(1), 119–133. <https://doi.org/10.1007/s10734-011-9484-1>
- Forsyth, R., Cullen, R., Ringan, N., & Stubbs, M. (2015). Supporting the development of assessment literacy of staff through institutional process change. *London Review of Education*, 13(3), 34. <https://doi.org/10.18546/LRE.13.3.05>
- Fox, J., & Weisberg, S. (2020). *car: Companion to Applied Regression*. [R package]. Retrieved from <https://cran.r-project.org/package=car>.
- Frey, B., & Overfield, K. (2002). Action research and professional development of faculty. *Canadian Journal for the Study of Adult Education*, 16(1), 20–33. Retrieved from <https://cjsae.library.dal.ca/index.php/cjsae/article/view/1894>
- Fuller, M.B., Skidmore, S.T., Bustamante, R.M., & Holzweiss, P.C. (2016). Empirically exploring higher education cultures of assessment. *The Review of Higher Education* 39(3), 395–429. [doi:10.1353/rhe.2016.0022](https://doi.org/10.1353/rhe.2016.0022).
- Fulmer, G., Lee, I., & Tan, K. (2015). Multi-level model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy, & Practice*, 1–20. <http://dx.doi.org/10.1080/0969594X.2015.1017445>
- Gilboy, M. B., Heinerichs, S., & Pazzaglia, G. (2015). Enhancing student engagement using the flipped classroom. *Journal of Nutrition Education and Behavior*, 47(1), 109–114. <https://doi.org/10.1016/j.jneb.2014.08.008>

- Goe, L., Bell, C., & Little, O. (2008). Approaches to evaluating teacher effectiveness: A research synthesis. In *National Comprehensive Center for Teacher Quality*. National Comprehensive Center for Teacher Quality. <https://eric.ed.gov/?id=ED521228>
- Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, 33(2), 14–18. <https://doi.org/10.1111/emip.12030>
- Goubeaud, K. (2010). How Is Science Learning Assessed at the Postsecondary Level? Assessment and Grading Practices in College Biology, Chemistry and Physics. *Journal of Science Education and Technology*, 19(3), 237–245.
- Grunwald, H., & Peterson, M. W. (2003). Factors that promote faculty involvement in and satisfaction with institutional and classroom student assessment. *Research in Higher Education*, 44(2), 173–204. <https://doi.org/10.1023/A:1022051728874>
- Guetterman, T. C., & Mitchell, N. (2016). The role of leadership and culture in creating meaningful assessment: A mixed methods case study. *Innovative Higher Education*, 41(1), 43–57. <http://dx.doi.org/10.1007/s10755-015-9330-y>
- Halinen, K., Ruohoniemi, M., Katajavuori, N., & Virtanen, V. (2014). Life science teachers' discourse on assessment: A valuable insight into the variable conceptions of assessment in higher education. *Journal of Biological Education*, 48(1), 16–22.
- Hanson, J. (2019). What is the distinction between assessment evaluation, measurement, and testing according to the concept of classroom assessment?. *eNotes Editorial*, 26 Sep. 2010, <https://www.enotes.com/homework-help/what-is-the-distinction-between-assessment-201131>. Accessed 15 Sep. 2020.
- Harland, T., McLean, A., Wass, R., Miller, E., & Sim, K. N. (2015). An assessment arms race and its fallout: High-stakes grading and the case for slow scholarship. *Assessment & Evaluation in Higher Education*, 40(4), 528–541. <https://doi.org/10.1080/02602938.2014.931927>
- Harris, L. R., & Brown, G. T. L. (2009). The complexity of teachers' conceptions of assessment: Tensions between the needs of schools and students. *Assessment in Education: Principles, Policy & Practice*, 16(3), 365–381. <https://doi.org/10.1080/09695940903319745>
- Haviland, D., Turley, S., & Shin, S.-H. (2011). Changes over time in faculty attitudes, confidence, and understanding as related to program assessment. *Issues in Teacher Education*, 20(1), 69–84. <https://eric.ed.gov/?id=EJ922283>
- Haviland, D., Shin, S., & Turley, S. (2010). Now I'm ready: The impact of a professional development initiative on faculty concerns with program assessment. *Innovative Higher Education; New York*, 35(4), 261–275. <http://dx.doi.org/10.1007/s10755-010-9140-1>
- Heinrich, W. F. (2015). *Exploring enacted mental models of learning outcomes assessment in higher education* [Ph.D., Michigan State University]. <http://search.proquest.com/docview/1734103555/abstract/65196ED5CCED4B04PQ/1>
- Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser,

- J., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, 76, 181–193. <https://doi.org/10.1016/j.tate.2017.12.001>
- Hidri, S. (2016). Conceptions of assessment: Investigating what assessment means to secondary and university teachers. *Arab Journal of Applied Linguistics*, 1(1), 19–43. <https://eric.ed.gov/?id=EJ1207924>
- Hines, S. R. (2009). Investigating Faculty Development Program Assessment Practices: What's Being Done and How Can It Be Improved? *The Journal of Faculty Development*, 23(3), 5–19. Education Database. <http://cyber.usask.ca/login?url=https://search.proquest.com/docview/868918040?accountid=14739>
- Hirschfeld, G. H. F., & Brown, G. T. L. (2009). Students' Conceptions of Assessment. *European Journal of Psychological Assessment*, 25(1), 30–38. <https://doi.org/10.1027/1015-5759.25.1.30>
- Hiser, C., & Francis, C. (2000). Intelligence: A brief history. *Journal of Inquiry and Research*, 72, 117 – 132. <http://doi.org/10.18956/00006382>
- Hodgson, C. (2008). *Assess — Podictionary Word of the Day*. OUPblog. Retrieved September 22, 2020 from <https://blog.oup.com/2008/05/assess/>
- Holroyd, C. (2000). Are assessors professional? Student assessment and the professionalism of academics. *Active Learning in Higher Education*, 1, 28–44. <https://doi.org/10.1177/1469787400001001003>
- Hootz, T., Mykota, D. B., & Fauchoux, L. (2016). Strengths-based crisis programming: Evaluating the process of care. *Evaluation and Program Planning*, 54, 50–62. <https://doi.org/cyber.usask.ca/10.1016/j.evalprogplan.2015.09.001>
- Huber, S. G., & Skedsmo, G. (2016). Teacher evaluation—Accountability and improving teaching practices. *Educational Assessment, Evaluation and Accountability*, 28(2), 105–109. <https://doi.org/10.1007/s11092-016-9241-1>
- Hughes, C. (2009). Framing the activities of institutions and academic development units in support of assessment. *Innovations in Education and Teaching International*, 46(2), 123–133.
- Hughes, G., & Hargreaves, E. (2015). Assessment literacy: Understanding relationships between feedback and learning. *London Review of Education*, 13(3), 1–2.
- Hutchings, P. (2011). From departmental to disciplinary assessment: Deepening faculty engagement. *Change*, 43(5), 36. Education Database. <http://cyber.usask.ca/login?url=https://search.proquest.com/docview/898695715?accountid=14739>
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385–402. <https://doi.org/10.1177/0265532208090158>

- Ion, G., & Cano, E. (2011). Assessment practices at Spanish universities: From a learning to a competencies approach. *Evaluation & Research in Education*, 24(3), 167.
<http://cyber.usask.ca/login?url=https://search.proquest.com/docview/895871183?accountid=14739>
- Janke, K. K., Kelley, K. A., Sweet, B. V., & Kuba, S. E. (2016). A modified Delphi process to define competencies for assessment leads supporting a Doctor of Pharmacy program. *American Journal of Pharmaceutical Education*, 80(10), 167.
<https://doi.org/10.5688/ajpe8010167>
- Jawitz, J. (2008). Learning to assess in the academic workplace: Case study in the natural sciences. *South African Journal of Higher Education*, 22(5), 1006–1018.
- Jawitz, J. (2009). Learning in the academic workplace: The harmonization of the collective and the individual habitus. *Studies in Higher Education*, 34(6), 601–614.
- Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing*, 30(3), 345–362.
<https://doi.org/10.1177/0265532213480334>
- Jurczyk, K., Vob, G., & Weihrich, M. (2016). Conduct of everyday life in subject-oriented sociology: Concept and empirical research. In E. Shraube & C. Højholt (Eds.), *Psychology and the Conduct of Everyday Life* (pp. 34–64). New York: Routledge.
- Kizlik, B. (2012). *Measurement, assessment, and evaluation in education*. Education information for new and future teachers. <http://www.adprima.com/measurement.htm>
- Kozma, R. B. (2003). ICT and educational change: A global phenomenon. In R. B. Kozma (Ed.), *Technology, innovation, and educational change: A global perspective* (pp. 1–18). Eugene, OR: International Society for Technology in Education.
- Kramer, P. I. (2009). The art of making assessment anti-venom: Injecting assessment in small doses to create a faculty culture of assessment. *Assessment Update*, 21(6), 8–10.
- Kremmel, B., & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100–120. Education Database.
<https://doi.org/10.1080/15434303.2019.1674855>
- Kvasova, O., & Kavytska, T. (2014). The assessment competence of university foreign language teachers: A Ukrainian perspective. *Language Learning in Higher Education*, 4(1), 159–177. <https://doi.org/10.1515/cercles-2014-0010>
- Landis, J., & Koch, G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 33(2), 363–374. doi:10.2307/2529786
- Leary, T. D., IV. (2017). *The lived faculty experience with formalized assessment initiatives: an interpretive phenomenological analysis* [Doctoral dissertation, Northeastern University]. Northeastern University Library Digital Repository Service.
<http://hdl.handle.net/2047/D20248508>

- Little, O., Goe, L., & Bell, C. (2009). A Practical Guide to Evaluating Teacher Effectiveness. In *National Comprehensive Center for Teacher Quality*. National Comprehensive Center for Teacher Quality. <https://eric.ed.gov/?id=ED543776>
- Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice*, 25(5), 442–467. <https://doi.org/10.1080/0969594X.2016.1268090>
- Macdonald, S. K., Williams, L. M., Lazowski, R. A., Horst, S. J., & Barron, K. E. (2014). Faculty attitudes toward general education assessment: A qualitative study about their motivation. *Research & Practice in Assessment; Lynchburg*, 9. <http://search.proquest.com/education/docview/1611488997/abstract/C861EEDEAB434C68PQ/2>
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329–344. <https://doi.org/10.1177/0265532213480129>
- Marrs, H. (2009). Perceptions of College Faculty Regarding Outcomes Assessment. *International Electronic Journal for Leadership in Learning*, 13(2). <https://eric.ed.gov/?id=EJ940619>
- McCune, V. (2018). Experienced academics’ pedagogical development in higher education: Time, technologies, and conversations. *Oxford Review of Education*, 44(3), 307–321. DOI: [10.1080/03054985.2017.1389712](https://doi.org/10.1080/03054985.2017.1389712)
- Medland, E. (2015). Examining the assessment literacy of external examiners. *London Review of Education; London*, 13(3), 21. <http://search.proquest.com/education/docview/1738011745/3F3CB02F890442C6PQ/2>
- Medland, E. (2016). Assessment in higher education: Drivers, barriers and directions for change in the UK. *Assessment & Evaluation in Higher Education*, 41(1), 81–96. <https://doi.org/10.1080/02602938.2014.982072>
- Medland, E. (2019). ‘I’m an assessment illiterate’: Towards a shared discourse of assessment literacy for external examiners. *Assessment & Evaluation in Higher Education*, 44(4), 565–580. <https://doi.org/10.1080/02602938.2018.1523363>
- Melguizo, T., Kosiewicz, H., Prather, G., & Bos, J. (2014). How are community college students assessed and placed in developmental math? Grounding our understanding in reality. *The Journal of Higher Education*, 85(5), 691. Education Database. <http://cyber.usask.ca/login?url=https://search.proquest.com/docview/1563679246?accountid=14739>
- Mertler, C. A. (2003, October). *Preservice versus inservice teachers’ assessment literacy: Does classroom experience make a difference?* Paper presented at the meeting of the Mid-Western Educational Research Association, Columbus, OH.
- Mertler, C. A., & Campbell, C. (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, Canada.

- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D.G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine* 6(7): e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Myers, S. M., & Myers, C. B. (2015). The confluence of family and academic lives: Implications for assessment practices of college student learning in higher education institutions in the United States. *The Social Science Journal*, 52(4), 527–535. <https://doi.org/10.1016/j.soscij.2015.03.003>
- Mykota, D. B., & Muhajarine, N. (2005). Community resilience impact on child and youth health outcomes: A neighbourhood case study. *Canadian Journal of School Psychology*, 20(1-2), 5–20. <https://doi-org.cyber.usask.ca/10.1177/0829573506295464>
- Norcini, J., Anderson, B., Bollela, V., Burch, V., Costa, M. J., Duvivier, R., Galbraith, R., Hays, R., Kent, A., Perrott, V., & Roberts, T. (2011). Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher*, 33(3), 206–214. <https://doi.org/10.3109/0142159X.2011.551559>
- Norton, L., Aiyegbayo, O., Harrington, K., Elander, J., & Reddy, P. (2010). New lecturers' beliefs about learning, teaching and assessment in higher education: The role of the PGCLTHE programme. *Innovations in Education and Teaching International*, 47(4), 345–356. <https://doi.org/10.1080/14703297.2010.518426>
- Norton, L., Floyd, S., & Norton, B. (2019). Lecturers' views of assessment design, marking and feedback in higher education: A case for professionalisation? *Assessment and Evaluation in Higher Education*, 44(8), 1209–1221. <https://doi.org/10.1080/02602938.2019.1592110>
- O'Loughlin, K. (2011). The interpretation and use of proficiency test scores in university selection: How valid and ethical are they? *Language Assessment Quarterly*, 8(2), 146–160. <https://doi.org/10.1080/15434303.2011.564698>
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing; London*, 30(3), 363–380. <http://dx.doi.org/10.1177/0265532213480336>
- O'Sullivan, R. G., & Johnson, R. L. (1993). *Using Performance Assessments To Measure Teachers' Competence in Classroom Assessment*. <https://eric.ed.gov/?id=ED358156>
- Ölmezer-Öztürk, E., & Aydin, B. (2018). Toward measuring language teachers' assessment knowledge: Development and validation of Language Assessment Knowledge Scale (LAKS). *Language Testing in Asia*, 8(1), 1–15. <https://doi.org/10.1186/s40468-018-0075-2>
- Overton, T. (2012). *Assessing learners with special needs: An applied approach*. Seventh Edition. Upper Saddle River, New Jersey: Pearson Education.
- Pawlyshyn, N. B. (2013). *Faculty engagement with higher education learning outcomes assessment: Transformation in a faculty learning community* [Ph.D., Capella University]. <http://search.proquest.com/education/docview/1431980485/abstract/5C1B79F2E1EA4A02PQ/67>
- Peters, M. D. J., Godfrey, C. M., Khalil, H., McInerney, P., Parker, D., & Soares, C. B. (2015). Guidance for conducting systematic scoping reviews. *JBI Evidence Implementation*, 13(3).

https://journals.lww.com/ijebh/Fulltext/2015/09000/Guidance_for_conducting_systematic_scoping_reviews.5.aspx

- Peterson, S. G. (2019). Epistemic beliefs and the innovation-decision process: A mixed methods analysis of faculty classroom assessment practice. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 80(6-A(E)).
- Pigott, T. D., Valentine, J. C., Polanin, J. R., Williams, R. T., & Canada, D. D. (2013). Outcome-Reporting Bias in Education Research. *Educational Researcher*, 42(8), 424–432. <https://doi.org/10.3102/0013189X13507104>
- Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10–12. <https://doi.org/10.1111/j.1745-3992.1993.tb00548.x>
- Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator’s confession. *The Teacher Educator*, 46(4), 265–273. <https://doi.org/10.1080/08878730.2011.605048>
- Presley, M. L. (2015). Investigating how participatory action research and the use of assessment instruments can support college instructors’ science assessment literacy (1850492560) [Doctoral dissertation, University of Missouri - Columbia]. ProQuest Dissertations & Theses Global. <http://cyber.usask.ca/login?url=https://search.proquest.com/docview/1850492560?accountid=14739>
- R Core Team (2020). *R: A Language and environment for statistical computing*. (Version 4.0) [Computer software]. Retrieved from <https://cran.r-project.org>. (R packages retrieved from MRAN snapshot 2020-08-24).
- Rawlasyk, P. (2016). *Exploring Assessment Practices in Higher Education: A Focus on Learning-oriented Assessment* [Ph.D., Northcentral University]. <http://search.proquest.com/education/docview/1809102071/abstract/5F862901901C495CPQ/7>
- Reder, M., & Crimmins, C. (2018). Why assessment and faculty development need each other: Notes on using evidence to improve student learning. *Research & Practice in Assessment*, 13. <http://cyber.usask.ca/login?url=https://search.proquest.com/docview/2191303535?accountid=14739>
- Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research*. [R package]. Retrieved from <https://cran.r-project.org/package=psych>.
- Richards, K., & Pilcher, N. (2014). Contextualising higher education assessment task words with an “anti-glossary” approach. *International Journal of Qualitative Studies in Education*, 27(5), 604–625. <https://doi.org/10.1080/09518398.2013.805443>
- Rogers, W. T. (1993). Principles for fair student assessment practices for education in Canada. *Canadian Journal of School Psychology*, 9(1), 110-127. <https://doi.org/10.1177/082957358500900111>

- Rorty, R. (2009). *Contingency, irony, and solidarity*. Cambridge, UK: Cambridge University Press.
- Rosa, M. J., Sarrico, C. S., & Amaral, A. (2012). Academics' perceptions on the purposes of quality assessment. *Quality in Higher Education*, 18(3), 349. Education Database. <http://cyber.usask.ca/login?url=https://search.proquest.com/docview/1112411594?accountid=14739>
- Sarrico, C.S., Rosa, M.J., Teixeira, P.N., & Cardoso, M. F. (2010) Assessing quality and evaluating performance in higher education: Worlds apart or complementary views?. *Minerva*, 48, 35–54 (2010). <https://doi.org/10.1007/s11024-010-9142-2>
- Schoepp, K., & Tezcan-Unal, B. (2017). Examining the effectiveness of a learning outcomes assessment program: a four frames perspective. *Innovative Higher Education*, 42(4), 305–319. <https://doi.org/10.1007/s10755-016-9384-5>
- Schraube, E., & Marvakis, A. (2016). Frozen fluidity: Digital technologies and the transformation of students' learning and conduct of everyday life. In E. Shraube & C. Højholt (Eds.), *Psychology and the Conduct of Everyday Life* (pp. 205-225). New York: Routledge.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (eds), *Perspectives of Curriculum Evaluation*. Chicago: Rand McNally, 39 – 83.
- Seema, R., Udam, M., & Mattisen, H. (2016). Attitudes of academic staff towards their own work and towards external evaluation, from the perspective of self-determination theory: Estonian case. *Quality in Higher Education*, 22(2), 117–126. <http://dx.doi.org/10.1080/13538322.2016.1195967>
- Skidmore, S. T., Hsu, H.-Y., & Fuller, M. (2018). A person-centred approach to understanding cultures of assessment. *Assessment and Evaluation in Higher Education*, 43(8), 1241–1257. <https://doi.org/10.1080/02602938.2018.1447082>
- Spearman, C. (1927). *The Abilities of Man*. London: Macmillan.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan* 72, 534–539.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77, 238–245.
- Swarat, S., Oliver, P. H., Tran, L., Childers, J. G., Tiwari, B., & Babcock, J. L. (2017). How Disciplinary Differences Shape Student Learning Outcome Assessment: A Case Study. *AERA Open*, 3(1).
- Taras, M. (2016). Assessment – Summative and formative – Some theoretical reflections. In M. Malloch, L. Cairns, & B. N. O'Connor (eds), *Learning and Teaching in Higher Education: Volume 4: Assessment and evaluation in higher education*. London: SAGE Publications, 466-78.
- Taras, M., & Davies, M. S. (2014). Perceptions and realities in assessment definitions and uses. *International Research in Education*, 2(1), 103–117. <https://doi.org/10.5296/ire.v2i1.4065>
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36. <https://doi.org/10.1017/S0267190509090035>

- Taylor, L. (2013). Communicating the theory, practice, and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412. <https://doi-org.cyber.usask.ca/10.1177/0265532213480338>
- The jamovi project (2021). *jamovi*. (Version 1.6) [Computer Software]. Retrieved from <https://www.jamovi.org>.
- Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2), 237-246. <https://doi.org/10.1177%2F1098214005283748>
- Tovar Klinger, C. J. (2016). *EFL professors' beliefs of assessment practices in an EFL pre-service teacher training undergraduate program in Colombia* [Ph.D., Southern Illinois University at Carbondale]. <http://search.proquest.com/education/docview/1871024514/abstract/5C1D779473524B1B/PQ/2>
- Tremblay, K., Lalancette, D., & Roseveare, D. (2012). *Assessment of higher education learning outcomes: AHELO feasibility study report. Volume 1 – Design and Implementation*. Retrieved from <http://www.oecd.org/education/highereducationandadultlearning/AHELOFSReportVolume1.pdf>
- Turner, C. C. (2013). *At cross purposes: Faculty perceptions of outcomes assessment and program change* [Doctoral Dissertation, Texas A&M University - Commerce]. Proquest Dissertations & Theses Global. <http://search.proquest.com/education/docview/1492361013/abstract/5C1B79F2E1EA4A02/PQ/27>
- Volante, L., & Ben Jaafar, S. (2008). Educational assessment in Canada. *Assessment in Education: Principles, Policy & Practice*, 15(2), 201–210. <https://doi.org/10.1080/09695940802164226>
- Wellman, J. V. (2010, January). *Connecting the dots between learning and resources*. (Occasional Paper No. 3). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA). Retrieved from <https://www.learningoutcomesassessment.org/wp-content/uploads/2019/02/OccasionalPaper3.pdf>
- Williams, B., Onsman, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, 8(3), Article 3. <https://doi.org/10.33151/ajp.8.3.93>
- Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, 40(2), 241–256. <https://doi.org/10.1007/s13384-013-0089-9>
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>

Appendix A: Thematic Summary

	Author & Date	Summary and Implications
Theme: Conceptualizing Assessment Literacy		
<i>General</i>		
1, 2	Medland, 2015, 2019	Identified key elements underpinning concept of AL
<i>Language assessment literacy</i>		
3	Inbar-Lourie, 2008	Components of a knowledge base specific to language assessment literacy
4	Ölmezer-Öztürk & Aydin, 2018	Core component is applied knowledge
5	Baker, Tsushima, & Wang, 2014	Defining language assessment literacy as a profile of competencies rather than a knowledge base
6	Kremmel & Harding, 2020	Empirically conceptualizes components of needed AL in language assessment contexts
7	Baker, 2016	Language assessment literacy as a profile of competencies, variable based on personal context
<i>Assessment Literacy for leadership</i>		
8, 9	Janke et al., 2016, 2019	Assessment-Literate leaders contribute to a culture of assessment that promotes assessment-literate faculty. More sophisticated assessment competencies necessary for leadership
<i>Classroom assessment literacy</i>		
10	White, 2008	Instructors as learners about their own classroom assessment theory and practices; AL to improve teaching and learning
Theme: Knowledge and Practices for Higher Education Assessment		
<i>Fluency in Assessment Language</i>		
11	Medland, 2016	AL necessary for shared understanding of assessment, but shared understanding is also necessary to be assessment literate
12	Baas, Rhoads, Thomas, 2016	Knowledge and Theory allows and enables dialogue, communication
13	Carless, 2009	AL facilitates trust which facilitates effective assessment. Sufficient knowledge to ensure others are using their knowledge appropriately

14, 15	Davies & Taras, 2016, 2018	Shared understandings about assessment are not addressed systemically and inclusively; they need reinventing regularly and within disciplinary communities. Contextual factors should come after understanding of theory, as theory remains stable across context. AL is needed to make informed and ethical decisions, to communicate about assessment processes or products created within any given context
16	Forsyth et al., 2015	Lack of clarity regarding assessment language produces difficulty with the grammar
17	Hanauer & Bauerle, 2015	vocabulary as an indicator of expertise/experience
18	Hawe, 2003	Need for institutional support in faculty understanding/comprehension of purposes of assessment to encourage appropriate use/participation
19	Keith & Brown, 2004	Clarifies language and purposes of assessment to provide a basic understanding
20, 21, 22	Raker, Emenike, & Holme, 2013; Emenike, Raker & Holme, 2013, Raker & Holme, 2014	Degrees in understanding of assessment terminology based on discipline. Personal understanding of assessment theory enables dialogue
23	Richards & Pilcher, 2014	Disciplinary use of assessment language. Contextual construction of assessment literacy
24, 25	Taras & Davies, 2013, 2017	ubiquitous understanding' of assessment language as basis for collaboration. Discipline-specific interpretations/valuations of terminology
<i>Knowledge Base for Assessment</i>		
26	Adachi et al., 2018	Appropriate understanding of techniques, awareness, and implementation of empirically supported techniques (scholarship)
27	Beebe, 2010	stability of assessment theory across contexts
28	Dietrich, 2011	Use of multiple methods to gain data, multiple sources of data to make judgements
29	McNeill et al., 2012	AL necessary to guide appropriate task design/selection, rather than being guided by perceptions
30	Melguizo et al., 2014	Practice without sufficient knowledge reinforces a cycle of inefficiency
31	Norcini et al., 2011	Disciplinary knowledge, awareness, and consistent implementation of 'best practices' for purpose in context
32	Taras & Davies, 2014	understanding of assessment that is clear, cogent, coherent, and shared; disciplinary variation in understanding
33	Taras, 2008	Understand definitions and terminology. Understand how processes relate to each other, and evaluate impact of our understanding on our practice

Disciplinary Knowledge

34	Fletcher et al., 2011	understanding of assessment practices bound to disciplinary content/processes. Knowledge base 'demystifies' assessment by making processes tangible, discussible
35	Goubeaud, 2010	disciplinary differences in assessment task selection
36	Harland et al., 2015	Variation of practices between disciplines
37	Hutchings, 2011	Anchor assessment within disciplines to promote faculty engagement
38	Jeong, 2013	What is drawn from the knowledge base is impacted by discipline, personal background
39	Malone, 2013	Contextual dependency on AL knowledge base selection. Disciplinary knowledge/purposes of assessment impacting AL knowledge use, but ALSO VALUE
40	Rawlusk, 2016	Significant differences in practices among disciplines
41	Swarat, Oliver, Tran, Childers, Tiwari, Babcock	Understanding disciplinary variation in assessment knowledge, which then shapes conceptions and practices.
42	Yeo & Boman, 2019	AL considered within disciplinary context. Universal approach to assessment [literacy] is not realistic

Understanding promotes participation

43	Bandy et al., 2016	Good' assessment occurs with understanding of higher-order purposes. critical reflexivity of assessment meanings. Self-regulation of assessor 'way of being'
44	Cohen, 2004	Knowledge of assessment theory allows for improved conceptions, implementation, and thus decision-making. Need for development to support depth and breadth of theoretical understanding
45	Cole & De Maio, 2009	AL may mediate reduced resistance to assessment: When faculty are unfamiliar with assessment, they may view it more negatively
46	Hines, 2009	Lack of knowledge created 'rote' engagement that was self-defeating/superficial
47	Macdonald et al., 2014	Resistance to assessment may occur when purposes are not understood. Current and future/expected assessment abilities. Necessity of support and opportunity for assessment development/practice
48	Marrs, 2009	Resistance seems to be associated with lack of understanding about 'what assessment is'
49	Pawlyshyn, 2013	AI informs engagement; engagement enables new approaches, which inform assessment at different levels
50	Tovar-Klinger, 2016	AL resists conception of assessment as taboo. Need for institutions to support AL of faculty

Standards based assessment/Outcomes Assessment

51	Alonzo et al., 2019	Conceptualizes standard-based assessment, finding six factors: writing outcomes, clarifying outcomes, using a range of strategies, scaffolding, ensuring trustworthiness of practices, analysing, and reflecting on data
52	Bahous & Nabhani, 2015	Effective assessment requires faculty with know-how, enthusiasm, humility; institutional support for faculty in these areas
53	Bloxham, 2012	Informed by a range of interacting processes: Development use of both explicit and implicit standards, engagement with students' work, discussing with internal/external colleagues about processes, incorporation of previous experience
54	BoarerPitchford, 2014	Selecting appropriate assessments. Personal demographics affect assessment literacy
55	Dahl, 2006	Individual bias affects assessment practice and may override group agreement. Faculty should be explicitly encultured TO MEANING AND USE OF CRITERIA
56	Hicks, 2016	Faculty should be explicitly encultured TO MEANING AND USE OF CRITERIA
57	Panadero et al., 2019	Lecturers using a variety of methods to assess students, but bound to 'traditional' assessment profiles

Student Assessment

58	Bennett et al., 2017	Designing assessment: Discern appropriate medium for assessment
59	Watkins et al., 2005	Understand how assessment feedback impacts learning

Knowledge of Stakeholders

60	Bresciani, 2012	clarity in audience for assessment data, indicators each audience requires, and intentions for audience data use
61	Seema et al., 2016	Greater investment/engagement in development improves management of external evaluation.

Theme: Conceptions of Assessment

Factors shaping conceptions

62	Alsobrook, 2010	Personal demographic variables impact construction of AL for a given individual (i.e., gender, ethnicity)
63	DiLoreto, 2013	Conceptions/ knowledge shaped by sufficient appropriate assessment experience. Could be inferred that inappropriate assessment experience may be difficult to undo

64	Dueben, 2015	Attitudes towards student assessment shaped by discipline
65	Ebersole, 2009	Faculty with more experience have higher levels of participation/unique attitudes towards assessment
66	Feuerstein, 2015	Epistemological compatibility with assessment processes. Comprehension of external assessment pressures/'bigger picture' of assessment.
67	Gilles et al., 2011	values and perceptions subject to contextual environmental variables (institutional selectivity, class size, level of education, disciplinary tradition, cultural differences)
68	Halinen et al., 2014	disciplinary diffs in assessment practices and use of results, epistemological approaches
69	Hidri, 2016	Variation shaped by disciplinary context/community
70	Myers & Myers, 2015	complexity of context: personal variables (gender, marital status) can affect assessment practice
71	Norton, Norton, & Shannon, 2013	Institution and discipline shaped <i>desirable practice and perceptions</i> . Assessment experience can lower perception of constraints
72	Quesada-Serra et al., 2016	demographic/disciplinary differences in perceptions of importance/competence
73	Raaper, 2016	demographic/disciplinary differences in perceptions of importance/competence
74	Sellbjer, 2017	"individual variation of interpretation of assessment task based on level of engagement. Contextual (personal, institutional) variation in meaning of assessment concepts. AL facilitates effective dialogue which facilitates effective practice/development of further AL
75	Walloch, 2006	Beliefs/confidence impacted by personal demographics (e.g., years spent teaching). Practice impacted by beliefs (i.e., educators use practices they value)
76	Webber, 2012	Demographic/disciplinary differences on selection of techniques
77	Wright, 2005	impact of discipline; faculty value discipline
Beliefs		
78	Birt, 2018	beliefs influence movement of knowledge to practice
79	Dew, 2016	faculty conception of assessment as for improvement of learning a motivator for faculty to engage with processes

80	Deygers & Malone, 2019	Conceptions affect how knowledge base is understood, interpreted, and in specific professional context. Beliefs/pragmatic considerations seem to override empiricism/policy considerations/knowledge of best practices
81	Evans, 2010	Engagement with outcomes assessment includes affective and behavioural components, as well as time and energy spent on assessment
82	Fostaty Young, 2012	Beliefs had a greater influence on assessment practice in context than characteristics of the institution did
83	Johnson, 2013	Moderated by program/discipline. Influences selection of classroom assessment practices/techniques.
84	Pastore & Pentassuglia, 2016	assessment conceptions are complex, hierarchical, multidimensional, and interrelated
85	Postareff et al., 2012	Orientations towards assessment that exist along a continuum, operate as a road map mediating thinking and practices
86	Rosa, Sarrico, & Amaral, 2012	Assessment for quality assurance, conducted for different purposes: 1. Communication about what is required of academics/departments. Therefore, assessment is an important means of strategy communication and implementation. 2. Motivation. (?) a framework for quality assessment that encompasses a balanced mix of criteria, given that these criteria influence the behaviour of faculty. 3. control, providing feedback to departments and academics so that measures can be taken 4. improvement: assessing what drives successful implementation 5. innovation
87	Samuelowicz & Bain, 2002	Orientations towards assessment: Coherent pattern of beliefs inferred from/grounded in assessment practices and explanations of those practices. Relationship between assessment thinking and practice. Assessment beliefs along a developmental continuum.
88	Schilling & Schilling, 2008	Variation in assessment epistemologies among disciplines
89	Wawiye, 2016	Moderated by program/discipline. Influences selection of classroom assessment practices/techniques.

Cognitions

90	Gingerich et al., 2014	Three distinct but not mutually exclusive perspectives on assessor cognition: assessor as trainable. Fallibility of human assessors, idiosyncratic expertise. Assessment values/biases make assessment judgement possible, and it will never be possible to eliminate those attributes which differentiate assessors
91	Beeseler Thompson, 2017	Approaches cognitions about assessment from a social-ecological approach

92	Deneen & Boud, 2014	Assessment knowledge is interpreted through collaboration, in context, through beliefs
93	Hackman, 2017	Perceived benefits of assessment. Comprehending different purposes of assessment without needing to polarize/prioritize those that aren't contextually relevant
94	Harrison et al., 2017	Personal and collective beliefs shaped design/implementation. Participants relied heavily on their own prior experiences of being assessed.
95	Offerdahl & Tomanek, 2011	assessment thinking moves along developmental continuum. Assessment practice not influenced only by thinking, but also by contextual factors
96	Reimann & Sadler, 2017	Development of AL tends to not happen without specific intervention, or else happens slowly and non-reflectively. Bi-directional relationship between assessment thinking and practice
97	Wang & Hurley, 2012	Perception of assessment as scholarly activity associated with belief that it is valuable, compelling, and rewarding

Conceptions limiting AL

98	Hoffman, 2015	Belief that assessment is important does not guarantee assessment practice. Confidence/perception of own competence
99	Rohrbacher, 2017	Use of data beyond reporting absent; lack of shared governance in process negatively influenced faculty engagement with assessment.

Theme: Contexts

100	Bloxham et al., 2016	Super complexity of assessment judgement. Influenced by subject discipline norms, access to knowledge filtered through beliefs
101	Boud et al., 2018	assessment viewed through a 'practice' lens assumes contextual factors
102	Castiglia & Turi, 2011	Miscommunication between institutional goals and faculty goals
103	Chong, 2009	context creates different conditions for assessment, requiring flexible approaches/application of knowledge
104	Day et al., 2019	Constraints based on context can override the knowledge base
105	Heinrich, 2015	Contextual influences on faculty assessment: Disciplinary training and socialization, env and cultural influences, incentives, accountability. Reconciling external/internal demands from different assessment levels

10 6	McCune, 2018	Multi-layered complexity of context (micro/macro/meso, departmental, institutional, subjective, material, individual, social, etc.)
10 7	Myers et al, 2015	al mediates influence of personal/institutional characteristics on actual assessment practice
<i>Professional Role</i>		
10 8	Albert et al., 2012	Assessment of scholarly research responds to discipline. Academics 'value' empiricism but are also biased by their value of empiricism.
10 9	Arreola et al., 2003	Faculty AL as a 'meta professional' skill: not an area of content expertise, yet required within the holistic professorial role
11 0	Blumenstein, 2015	Professional influences (discipline specific interpretation). A collaborative approach to assessment skills development. Need for preparedness to communicate with others who have different levels of AL
11 1	Ewell et al., 2011	Disciplinary diffs in assessment practices and use of results. Different 'drivers' for assessment, use of results. AL benefits from direct support. Necessary AL varies by context
11 2	Harman & McDowell	Ongoing renegotiation of assessment continual discourse/identity
11 3	Holryod, 2000	Participation as a member of a communicating network allows for shared meaning of standards implicit in assessment criteria
11 4	Ion & Canes, 2011	Disciplinary impact on practice preference
11 5	Norton et al., 2010	Engagement with desirable practices related to professional knowledge (discipline, pedagogy, scholarship, etc.)
11 6	Norton, Floyd, & Norton, 2019	Engagement with desirable practices in assessment, underpinned by professional knowledge (e.g., discipline)
11 7	O'Loughlin, 2011	Language Assessment Literacy necessary for higher education professionals to make correct interpretations and just decisions
<i>Cultures of assessment</i>		
11 8	Emil, 2011	congruence or compatibility of personal beliefs/knowledge (al?) with local assessment culture
11 9	Fuller et al., 2016	Conceptualizes cultures of assessment, including Faculty Perceptions, Use of Data, Sharing, Compliance or Fear Motivators, and Normative Purpose of assessment

12 0	Hughes, 2010	Low AL identified in poor assessment culture
12 1	Skidmore, Hsu, Fuller, 2018	Examines different groups of faculty perceptions and behaviours regarding assessment

Theme: Development

12 2	Peterson, 2019	professional development improves sophistication of (classroom) assessment beliefs
12 3	Turner, 2013	Assessment Literacy develops along a continuum; knowledge base and conceptions move from less to more sophisticated

Institutional Support for Faculty AL Development

12 4	Beckwith et al., 2010	Institutional resources develop faculty AL; ownership and engagement facilitates and requires AL
12 5	Creason, 2015	Need for institutional resources/support for faculty to develop literacy, practice assessment. Elements needed to implement outcomes assessment: knowledge of SLO, clear plan, training, expertise, staff to assist faculty, time, department culture, faculty characteristics
12 6	Emil & Cress, 2014	Institutions support participation which can promote AL. Need for congruence between personal skills/values and organisation
12 7	Frey & Overfield, 2002	Need for resourced development opportunities
12 8	Grunwald & Peterson, 2003	Institutional support and development fortifies faculty satisfaction with assessment efforts, girding their participation
12 9	Guetterman & Mitchell, 2016	effective assessment requires engaged faculty, which requires them to be AL, which can be facilitated by institutional investment in faculty development
13 0	Haviland et al., 2011	Institutional support for professional development, promote faculty ownership, and present a clear and sustained message may elicit greater faculty engagement with assessment processes
13 1	Haviland, 2009	Inquiry, owned by faculty, over accreditation as driver for assessment. Faculty must become ultimate leaders over their assessment efforts
13 2	Kramer, 2008	Institutional support of faculty planning, designing, implementing assessment mitigates resistance, promotes literacy and culture of assessment
13 3	Reder & Crimmins, 2018	Partnership between faculty and developers enables use of data for faculty development, empowers faculty ownership

Literacy develops in context in practice

13 4, 13 5	Jawitz, 2008, 2009	AL develops with experiences. Assessment community membership links past experiences with future possibilities. Confidence achieved via repetition
13 6	Leary, 2017	assessment for improvement of own practice
13 7	Presley 2015	Engaging in PAR and using assessment instruments improved AL by developing confidence in beliefs, knowledge of strategies/interpretation, and integration of knowledge in ways that support changes in practice
13 8	Clark & Filinson, 2011	Faculty develop Assessment Literacy through engagement and appropriate assessment experience. Developing knowledge base facilitates assessment ownership.
13 9	Deeley & Bovill, 2017	improve AL through engagement in 'observation, imitation, dialogue, and practice'
14 0	Caudle, 2014	involves faculty involvement which I facilitated by opportunities for contextually appropriate involvement. Involves personal acceptance of assessment as a mechanism for improvement

Need for development of AL in HE

14 1	Brinkman-Staneva, 2015	Need for continual, explicit development of assessment knowledge, purposes, design, and implications
14 2	DiBiase-Lubrano. 2018	Need for targeted development of Language AL among higher education language assessment professionals
14 3	Elshawa et al., 2017	Need for sustained development opportunities to empower higher education instructors in their professorial assessment role
14 4	Kvasova & Kavytska, 2014	Need for continual, explicit development of assessment knowledge, purposes, design, and implications
14 5	O'Loughlin, 2013	Need for those who use language proficiency tests to be literate in the use and interpretation of such tests
14 6	Rodríguez-Gómez, Quesada-Sera & Ibarra-Sáiz, 2016	Need for continual, explicit development of assessment knowledge, purposes, design, and implications
14 7	Shams & Iqbal, 2019	Need for development of classroom assessment literacy of university teachers
14 8	Stencil, 2014	Indicated a lack of knowledge about assessment and a need for targeted development

Theme: Decision-making

Compromise of competing demands in context

14 9	Bearman et al., 2016	Different capacity to change assessments in different contexts. Assessment as a series of contextual assessor decisions influenced by disciplinary traditions
15 0	Dawson et al., 2013	Super complexity of context subsumes level, purposes, persons involved, as well as their personal status and capacity to make changes
15 1	Delaney, 2015	Assessment is continually renegotiated; AL involves participation in negotiation (and all that includes)
15 2	Green, 2016	make best decisions possible using a collaborative approach and strategic process to align all elements of an assessment cycle while considering the influences of internal and external factors
15 3, 15 4	Orr, 2011; Orr & Bloxham, 2013	Values embodied in personal, disciplinary narratives about assessment. Assessment decision-making occurs within webs of influence and experience
15 5	Stefaniak et al., 2018	form an opinion based on constrained information/resources that consider systemic factors influencing an environmental context

Experience and Expertise

15 6	Aloi, 2004	Faculty Assessment Literacy as an integral part of institutional decision-making processes
15 7	Büyükkarci, 2016	experience with assessment as opposed to experience as an academic
15 8	Filetti et al., 2010	with experience comes rigour
15 9	Haviland et al., 2010	faculty engage when they understand and value the purpose, use the data, perceive organizational support, and believe leaders value assessment. Ownership facilitates engagement and thus development of AL
16 0	Neushel & Rego, 2018	Designing own goals and strategies, discussing with others
16 1	Streff, 2016	decisions made by experienced assessors seem subconscious but influenced by learning objectives. experienced assessors continually revise their assessments

Using Data

16 2	Campbell, 2014	Data used in response to institutional planning. Faculty viewed as a means of data production
---------	----------------	---

16 3	Jonson et al., 2017	Use of assessment data for decision-making increases when evidence is action oriented, viewed as high quality, and when faculty are knowledgeable, have positive disposition towards assessment, and perceive institutional support.
---------	---------------------	--

16 4	Schoepp & Tezcan-Unal, 2017	To fully participate in assessment processes faculty must understand purposes, have their human needs addressed, share in process governance, and be sufficiently motivated
---------	-----------------------------	---

Judgement of evidence

16 5	Liu, 2008	al underscores rationale for selecting elements for a particular context
---------	-----------	--

16 6	Pool, et al., 2018	Arriving at same judgements despite differences in reasoning/interpretation of evidence due to divergent beliefs, performance theories, and inferences. Assessors should receive training providing insight into the factors affecting their own judgements.
---------	--------------------	--

Theme: Metacognition

16 7	Baker, 2014	Faculty perceptions/and use of student evaluation data to improve their own practice
---------	-------------	--

16 8	Fuller & Skidmore, 2014	congruence between personal values/beliefs and assessment cultures values/beliefs
---------	-------------------------	---

16 9	Lees & Anderson, 2015	Using assessment to improve your own practice. Alignment with practice and competency (e.g., 'practicing what you preach')
---------	-----------------------	--

17 0	Raubenheimer, 2004	Consistency between conceptions and practices, influence/alignment with beliefs about teaching/learning
---------	--------------------	---

17 1	Sadler & Reimann, 2018	Development tends to not happen without specific intervention, or else happens slowly and non-reflectively. Bi-directional relationship between assessment thinking and practice
---------	------------------------	--

17 2	Tlali & Jacobs, 2015	Consistency between conceptions and practices
---------	----------------------	---

Theme: Various Influencing Factors

Institutional Characteristics

17 3	Hughes, 2009	Assessment Literacy is institutionally supported by clarification of roles and practices. Faculty awareness of external pressures
---------	--------------	---

17 4	McCullough, 2008	Identified institutional characteristics pertaining to faculty satisfaction with assessment: ownership/participation, benefits to assessment, and assessment leadership
---------	------------------	---

17 5	Overton, 2010	Personal congruence with institutional assessment activities. Senior-level support for faculty-level assessment
---------	---------------	---

17 Scott, 2013 Assessment needs to account for power, economics of staffing, and differing ways of thinking

6

17 Talley, 2016 Institutional pressure for faculty to inflate grades

7

17 Tarakegne, 2019 Institutional commitment to assessment and provision of resources for assessment

8

Ownership

17 Colina De ViVero Need for faculty to feel ownership, valued in assessment process

9

18 Eyres, 2016 Ownership and engagement with local context

0

18 Piascik & Bird, 2008 Buy in at all levels, including support from administration, department chairs. Development essential to create/sustain culture of assessment. Faculty orientation revealing processes, roles, epistemologies is necessary

1

Culture of assessment

18 LaManna, 2019 Need for faculty to value process of assessment and participate in local culture of assessment

2

18 Nodye & Parker, 2010 Institutional support of CoA supports faculty ownership of assessment

3

Appendix B: Themes of Assessment Professionalism

Appendix C: Sources about HE Assessment Professionalism

	Author & Date	Summary and Implications
<i>External Evaluation</i>		
1	Rosa, Sarrico, & Amaral, 2012	Assessment for quality assurance, conducted for different purposes: 1. Communication about what is required of academics/departments. Therefore, assessment is an important means of strategy communication and implementation. 2. Motivation. (?) a framework for quality assessment that encompasses a balanced mix of criteria, given that these criteria influence the behaviour of faculty. 3. control, providing feedback to departments and academics so that measures can be taken 4. improvement: assessing what drives successful implementation 5. innovation
2, 3	Medland, 2015, 2019	Identified key elements underpinning concept of AL
4	Medland, 2016	AL necessary for shared understanding of assessment, but shared understanding is also necessary to be assessment literate
5	Seema et al., 2016	Greater investment/engagement in development improves management of external evaluation. Competent and intrinsically motivated academics perceived external evaluation more positively
<i>Assessing language skills and using data from language assessments for admission, hiring</i>		
6	Baker, 2016	Language assessment literacy as a profile of competencies, variable based on personal context
7	Baker, Tsushima, & Wang, 2014	Defining language assessment literacy as a profile of competencies rather than a knowledge base
8	Deygers & Malone, 2019	Conceptions affect how knowledge base is understood, interpreted, and in specific professional context. Beliefs/pragmatic considerations seem to override empiricism/policy considerations/knowledge of best practices
9	DiBiase-Lubrano, 2018	Need for targeted development of Language AL among higher education language assessment professionals
10	Inbar-Lourie, 2008	Components of a knowledge base specific to language assessment literacy
11	Jeong, 2013	What is drawn from the knowledge base is impacted by discipline, personal background
12	Kremmel & Harding, 2020	Empirically conceptualizes components of needed AL in language assessment contexts
13	Kvasova & Kavytska, 2014	Need for continual, explicit development of assessment knowledge, purposes, design, and implications
14	Malone, 2013	Contextually dependency on AL knowledge base selection. Disciplinary knowledge/purposes of assessment impacting AL knowledge use, but ALSO VALUE
15	O'Loughlin, 2011	Language Assessment Literacy necessary for higher education professionals to make correct interpretations and just decisions
16	O'Loughlin, 2013	Need for those who use language proficiency tests to be literate in the use and interpretation of such tests

17	Ölmezer-Öztürk & Aydin, 2018	Developed scale to measure language teachers LAL; Core component is applied knowledge
<i>Assessment for academic developers</i>		
18	Beckwith et al., 2010	Institutional resources develop faculty AL; ownership and engagement facilitates and requires AL
19	Hughes, 2009	Assessment Literacy is institutionally supported by clarification of roles and practices. Faculty awareness of external pressures
20	Reder & Crimmins, 2018	Partnership between faculty and developers enables use of data for faculty development, empowers faculty ownership
<i>Professional Roles</i>		
21	Albert et al., 2012	Assessment of scholarly research responds to discipline. Academics 'value' empiricism but are also biased by their value of empiricism.
22	Arreola et al., 2003	Faculty AL as a 'meta professional' skill: not an area of content expertise, yet required within the holistic professorial role
23	Bandy et al., 2016	Good' assessment occurs with understanding of higher-order purposes. critical reflexivity of assessment meanings. Self-regulation of assessor 'way of being'
24	Blumenstein, 2015	Professional influences (discipline specific interpretation). A collaborative approach to assessment skills development. Need for preparedness to communicate with others who have different levels of AL
25	Ewell et al., 2011	Disciplinary diffs in assessment practices and use of results. Different 'drivers' for assessment, use of results. AL benefits from direct support. Necessary AL varies by context
26	Feuerstein, 2015	Epistemological compatibility with assessment processes. Comprehension of external assessment pressures/'bigger picture' of assessment.
27	Harman & McDowell	Ongoing renegotiation of assessment continual discourse/identity
28	Holryod, 2000	Participation as a member of a communicating network allows for shared meaning of standards implicit in assessment criteria
29	Ion & Canes, 2011	Disciplinary impact on practice preference
30, 31	Janke et al., 2016, 2019	Assessment-Literate leaders contribute to a culture of assessment that promotes assessment-literate faculty. More sophisticated assessment competencies necessary for leadership
32	Norton et al., 2010	Engagement with desirable practices related to professional knowledge (discipline, pedagogy, scholarship, etc.)
33	Norton, Floyd, & Norton, 2019	Engagement with desirable practices in assessment, underpinned by professional knowledge (e.g., discipline)

<i>program assessment</i>		
34	Hines, 2009	Lack of knowledge created 'rote' engagement that was self-defeating/superficial
35	Haviland et al., 2010	faculty engage with program assessment when they understand and value the purpose, use the data, perceive organizational support, and believe leaders value assessment. Ownership facilitates engagement and thus development of AL
36	Emil & Cress, 2014	Institutions support participation which can promote AL. Need for congruence between personal skills/values and organisation
37	Emil, 2011	Congruence or compatibility of personal beliefs/knowledge (al?) with local assessment culture
<i>contribution to culture (?)</i>		
38	Fuller et al., 2016	Conceptualizes cultures of assessment, including Faculty Perceptions, Use of Data, Sharing, Compliance or Fear Motivators, and Normative Purpose of assessment
39	Hughes, 2010	Low AL identified in poor assessment culture
40	Skidmore, Hsu, Fuller, 2018	Examines different groups of faculty perceptions and behaviours regarding assessment
<i>admin/institutional frame</i>		
41	Campbell, 2014	Data used in response to institutional planning. Faculty viewed as a means of data production
<i>evaluation of faculty</i>		
42	Baker, 2014	Faculty perceptions/and use of student evaluation data to improve their own practice

Appendix C: Certificate of Ethical Approval



UNIVERSITY OF
SASKATCHEWAN

Behavioural Research Ethics Board (Beh-REB) 15-Dec-2020

Certificate of Approval

Application ID: 2347

Principal Investigator: Amin Mousavi

Department: Department of Educational Psychology
and Special Education

Locations Where Research

Activities are Conducted: Saskatchewan, Canada

Student(s): Derek Friesen

Funder(s): Office of the Vice-President Research

Sponsor: University of Saskatchewan

Title: Towards a Comprehensive View of Assessment Literacy for Higher Education Settings

Approved On: 15-Dec-2020

Expiry Date: 15-Dec-2021

Approval Of: Behavioural Research Ethics Application

Consent Forms (Survey and Item Review)

Item Review Contact Plan

Survey Contact Plan

Item Review Form

Acknowledgment Of: TCPS2 Core Certificate (Friesen)

Review Type: Delegated Review

CERTIFICATION

The University of Saskatchewan Behavioural Research Ethics Board (Beh-REB) is constituted and operates in accordance with the current version of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TPCS 2 2018). The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this project, and for ensuring that the authorized project is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol or consent process or documents.

Any significant changes to your proposed method, or your consent and recruitment procedures should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.

ONGOING REVIEW REQUIREMENTS

In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month prior to the current expiry date each year the project remains open, and upon project completion. Please refer to the following website for further instructions: <https://vpresearch.usask.ca/researchers/forms.php>.

Digitally Approved by Stephanie Martin
Vice-Chair, Behavioural Research Ethics Board
University of Saskatchewan

Appendix D: Survey Items

1. Landing Page and Consent

Thank you for your interest in participating in this survey regarding assessment literacy in higher education. Please read the survey information below and indicate your provision of informed consent.

Participant Consent Form

You are invited to participate in a research study entitled: Towards a Comprehensive View of Assessment Literacy for Higher Education Settings

Student Researcher: Derek Friesen, Graduate Student, Educational Psychology and Special Education, University of Saskatchewan, Derek.W.Friesen@usask.ca

Principal Investigator/Supervisor: Dr. Amin Mousavi, Assistant Professor, Educational Psychology and Special Education, University of Saskatchewan, Amin.Mousavi@usask.ca, 306-966-7653

Purpose and Objective of the Research:

We are doing a study to learn about how the concept of assessment literacy (AL) pertains to contexts in higher education (HE) settings. Our main purpose is to distinguish features of HE settings necessitating an AL more sophisticated than can be described with current AL models, given how such models have largely been developed through investigation of teachers in school-based settings. In phase one of this project, an empirical scoping review of studies investigating assessment literacy resulted in a dataset of 182 research sources relating to knowledge, skills, and conceptions of assessment necessary for effective assessment in HE settings. We identified a small but notable body of literature related to student learning assessment literacy, but a paucity of literature related to non-student assessment literacy tasks in higher education. Themes emerging from these bodies of literature were analyzed in comparison to existing AL conceptualization and discussed considering prior AL research.

For phase two, we used existing assessment literature and expert review to develop a questionnaire to assess the AL of faculty members in Canadian HE institutions. This survey targets faculty members in several Canadian HE institutes to provide baseline data for faculty understanding, perception, and involvement with tasks related to assessment literacy higher education in Canada. This an important step towards defining a unique framework of AL in HE which will allow for further enhancement of assessment practices in HE and provide a structure for targeted professional development for new and current faculty members in Canadian universities.

Procedures:

If you decide to take part in this study, you will participate by completing a survey about assessment literacy in higher education using the online survey platform SurveyMonkey. The survey consists of 50 items. Participants will respond to items using a Likert-type scale (i.e., 1 - 5) and a dichotomous scale (i.e., true or false). There are also two open-ended questions. Participants can complete the survey using their own computer. **Completing the survey will require an estimated 10-15 minutes.** Please feel free to contact the investigators with any questions regarding the procedures and goals of the study or your role.

Funded by:

This project was funded by a SSHRC grant.

Potential Risks:

There are no known or anticipated risks to you by participating in this research.

Potential Benefits:

You may derive personal satisfaction from contributing to research about assessment literacy in higher education settings.

Confidentiality:

Collected data will be used for statistical analysis. Results of the survey will be reported on in scientific meetings, journal articles, and graduate theses. The information collected from you will be aggregated with other participants in all reports. No identifying information will ever be revealed in reporting of results. Moreover, the consent forms will be stored separately from the data so that it will not be possible to associate a name with any given set of responses. This survey is hosted by Survey Monkey. Your data will be stored in facilities hosted in Canada. Please see the following for more information on Survey Monkey's Privacy Policy.

Storage of Data:

Survey data will be stored using Survey Monkey. Additionally, a copy of the raw data will securely stored by the principal investigator. During analysis, electronic data will be stored using a password-protected cloud-based storage product (i.e. OneDrive). After analysis, data will be stored for 5 years post-publication as per University of Saskatchewan guidelines. All data will be deleted by the PI after project completion and the minimum storage period.

Right to Withdraw:

You can participate or not. You can stop at any time. Your participation is voluntary, and you can stop the survey at any point until submitting the completed survey. Because the survey is anonymous, data withdrawal will not be possible after you submit the completed survey. Data collection will resolve by May 2021, after which data analysis will commence.

Follow up:

To obtain results from the study, please contact the principal investigator. Results will be published in a graduate thesis as well as submitted to relevant educational assessment and measurement journals for peer review. If you wish to be notified directly about published results, please notify the researcher in your response. It is expected that results will be available by August 2021.

Questions or Concerns:

Contact the researcher(s) using the information at the top of page 1. This research project has been approved on ethical grounds by the University of Saskatchewan Behavioural Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office: ethics.office@usask.ca; 306-966-2975; out of town participants may call toll free 1-888-966-2975.

By completing and submitting this questionnaire, your **free and informed consent** is implied and indicates that you understand the above conditions of participation in this study.

* 1. After reading the above information, please indicate your consent to participate.

- ☐ Yes, I have read the above information and consent to participate
- ☐ No, I do not consent to participate (survey will end)

2. Demographics

Please provide some basic information about yourself.

* 2. What is your age?

- ☐ Less than 25
- ☐ 25 to 34
- ☐ 35 to 44
- ☐ 45 to 54
- ☐ 55 to 64
- ☐ 65 to 74
- ☐ 75 or older

* 3. What is your gender?

- ☐ Woman (including transgender women)
- ☐ Man (including transgender men)
- ☐ Prefer not to say
- ☐ Prefer to self-describe as:

* 4. About how many years have you been employed as a faculty member in a higher education institution?

- ☐ Less than 5 years
- ☐ At least 5 year but less than 10 years
- ☐ At least 10 years but less than 15 years
- ☐ At least 15 years but less than 20 years
- ☐ At least 20 years but less than 25 years
- ☐ 25 years or more

* 5. What is your rank at your institution?

- ☐ Instructor
- ☐ Lecturer
- ☐ Assistant professor
- ☐ Other (please specify)
- ☐ Associate professor
- ☐ Full professor

* 6. Which of the following best describes your field of research and teaching?

- ☐ Mathematics
- ☐ Science
- ☐ Healthcare
- ☐ Medicine
- ☐ Computing
- ☐ Engineering
- ☐ Technology
- ☐ Business
- ☐ Education
- ☐ Other (please specify)

3. Assessment use in higher education

Thank you for your participation. This survey has been designed to investigate faculty understanding, perception of, and involvement with tasks related to assessment literacy higher education in Canada. Our priority is to distinguish features of higher education settings necessitating an AL more sophisticated than can be described with current AL models, given how such models have largely been developed through investigation of teachers in school-based settings.

To begin, please consider the following definition:

"In higher education, the term assessment refers to the process of gathering, documenting, analyzing, and interpreting empirical evidence to make decisions related to situations like (but not limited to) assessing student learning, program evaluation, admission and promotion."

With this definition in mind, please rate the following items.

* 7. How **frequently** is assessment involved/used for each of the professional roles or tasks in higher education below?

	Not at all	Occasionally	Sometimes	Often	All the time	N/A
Assessing student learning within courses	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluating course materials such as textbooks or research artifacts	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluating instruments for assessing student learning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Critiquing assessment tools	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Student admissions decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Engaging in feedback based on assessments	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

* 8. How **frequently** is assessment involved/used for each of the professional roles or tasks in higher education below?

	Not at all	Occasionally	Sometimes	Often	All the time	N/A
Using assessments to evaluate program outcomes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Using assessments to evaluate academic programs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Advising students using different types of assessments	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluating faculty research products/artifacts	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Evaluating theses/dissertations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

* 9. How **frequently** is assessment involved/used for each of the professional roles or tasks in higher education below?

	Not at all	Occasionally	Sometimes	Often	All the time	N/A
Faculty evaluations based on peer assessment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Faculty evaluations based on student evaluations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Faculty hiring decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Faculty tenure/promotion decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Leading departmental assessment efforts	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Participating in institutional assessment initiatives	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Research project management	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

4. True or false

Read each of the following statements and indicate if you perceive it as true or false. If the statement does not apply to you, indicate it as N/A (not applicable).

* 10. Please indicate if you believe each of the following statements to be **true** or **false**.

	True	False	N/A
Assessment processes result in information that is useful to me	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The same assessment principles underlie any assessment task	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Depending on who I am assessing, my interpretations could be different	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I can put my assessment knowledge into practice for any given task	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have a practical working knowledge of assessment theory	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I recognize professional situations where I can apply my assessment knowledge	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Similar assessment concepts can be applied to different situations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

* 11. Please indicate if you believe each of the following statements to be **true** or **false**.

	True	False	N/A
I reflect on my assessment knowledge and practices	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My opinion of assessment is shaped by my knowledge of assessment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I regard assessment as useful in my profession	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My assessment efforts are generally worthwhile	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My views of assessment have been shaped by experiences	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

* 12. Please indicate if you believe each of the following statements to be **true** or **false**.

	True	False	N/A
I am experienced at using assessment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I consider myself to be an assessor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I do not feel confident using assessment for tasks other than classroom assessment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I do not need to practice assessment for making decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel confident about when and how to use assessment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5. True or false

Read each of the following statements and indicate if you perceive it as true or false. If the statement does not apply to you, indicate it as N/A (not applicable).

* 13. Please indicate if you believe each of the following statements to be **true** or **false**.

	True	False	N/A
I clarify the meaning of the assessment concepts I use when discussing assessment with others	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I know how to communicate about assessment results and decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I understand the concepts other people use when discussing assessment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I use the term assessment in the same way as colleagues from other departments/colleges	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I use the term assessment in the same way as colleagues in my department/college	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

* 14. Please indicate if you believe each of the following statements to be **true** or **false**.

	True	False	N/A
I use assessment when I make professional decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In higher education, decision making is not possible without assessment skills	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I generally do not need assessment skills to perform my tasks	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I participate in community learning activities regarding assessment	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I seek out opportunities to increase my assessment literacy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am engaged with assessment processes in my profession	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

6. Open Ended items

Παραλλήλες δεξιότητες προαδεδεγμένες:

Ηνχηγερδύχαιον,ζηεζερεμ ασεσσεμ εντρεφεροεοζηεετοχεσσεφάατηερινγ, δοχυμ εντινγ,ανάλυζινγ,ανδεντερετινγ εμ περιχάλεπιδενχεετομ ακεδεχισιονσρελατεδεοτυατιονσλικε(βυτεοτλμ ιτεδεο)ασεσσινγ ετυδεναρνινγ,ετρογραμεπαλατιον,αδμ ισσιονανδεερομ οτιον.ΥΩιτηεηισδεφνιτιονενμ ινδ,χονσιδερεζηεφολωινγ θυεσιονσ:

15. Other than classroom assessment, how is assessment commonly used in higher education?

16. What do higher educators need to know and/or be able to do for these assessment tasks?

7. End of survey

Τησδουρεψιδεχομ πλετε.Τηανκβουφωρβουρεαρηχιατιον.

Appendix E: Frequency of Assessment in HE Tasks Correlation Table

Correlation Matrix

	item1	item2	item3	item4	item5	item6	item7	item8	item9	item10	item11	item12	item13	item14	item15	item16	item17	item18
item1	—																	
item2	0.577 ***	—																
item3	0.473 ***	0.594 ***	—															
item4	0.408 **	0.634 ***	0.789 ***	—														
item5	0.289 *	0.079	0.150	0.086	—													
item6	0.451 ***	0.394 **	0.742 ***	0.677 ***	0.082	—												
item7	0.603 ***	0.498 ***	0.419 **	0.507 ***	0.286 *	0.437 ***	—											
item8	0.550 ***	0.451 ***	0.434 ***	0.508 ***	0.317 *	0.429 **	0.862 ***	—										
item9	0.421 **	0.419 **	0.434 ***	0.318 *	0.230	0.508 ***	0.568 ***	0.471 ***	—									
item10	0.494 ***	0.357 **	0.316 *	0.388 **	0.363 **	0.131	0.482 ***	0.451 ***	0.344 *	—								
item11	0.565 ***	0.328 *	0.192	0.230	0.461 ***	0.215	0.440 ***	0.499 ***	0.417 **	0.662 ***	—							
item12	0.560 ***	0.345 **	0.338 *	0.304 *	0.399 **	0.238	0.358 **	0.341 *	0.362 **	0.707 ***	0.615 ***	—						
item13	0.450 ***	0.175	0.048	0.043	0.280 *	0.009	0.327 *	0.288 *	0.291 *	0.438 ***	0.441 ***	0.395 **	—					
item14	0.481 ***	0.332 *	0.239	0.295 *	0.450 ***	0.111	0.309 *	0.220	0.276 *	0.724 ***	0.482 ***	0.639 ***	0.390 **	—				
item15	0.562 ***	0.347 **	0.307 *	0.242	0.390 **	0.191	0.312 *	0.337 *	0.334 *	0.705 ***	0.782 ***	0.723 ***	0.465 ***	0.724 ***	—			
item16	0.391 **	0.378 **	0.414 **	0.326 *	0.258	0.453 ***	0.221	0.257	0.285 *	0.313 *	0.350 **	0.366 **	0.210	0.405 **	0.521 ***	—		
item17	0.390 **	0.229	0.408 **	0.283 *	0.572 ***	0.403 **	0.302 *	0.340 *	0.377 **	0.306 *	0.536 ***	0.354 **	0.277 *	0.385 **	0.518 ***	0.707 ***	—	
item18	0.405 **	0.374 **	0.287 *	0.374 **	0.298 *	0.264	0.434 ***	0.324 *	0.421 **	0.541 ***	0.519 ***	0.457 ***	0.330 *	0.529 ***	0.485 ***	0.318 *		

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Appendix F: Item Review Recruitment Materials and Consent Form

Hello Dr. [name],

My name is Derek Friesen and I am undertaking a research project for my graduate program in Measurement and Evaluation in the Educational Psychology and Special Education department at The University of Saskatchewan. For this project, I am developing a questionnaire to investigate higher educators' conceptualizations of assessment literacy. Having been influenced by your extensive work developing the concept of assessment literacy, I am inviting you to review the proposed item pool for its validity and content alignment with the sincere belief that your feedback will improve the quality of the instrument. The review will take about 30-40 minutes of your time. I have attached an information and consent form for your review. Please return this form with a signature (digital signature is fine) should you agree to participate. Upon your consent to review this instrument, I will provide the item review form including a proposed item pool and instructions for review. I would appreciate you completing the review by the 30th of October, with the possibility for an additional week of extension.

If you have any concerns and questions about this study, please feel free to contact me at derek.w.friesen@usask.ca. Thank you for your consideration.

Kind regards,

Derek Friesen

Item Review Participant Consent Form

You are invited to participate in a research study entitled: Towards a Comprehensive View of Assessment Literacy for Higher Education Settings

Student Researcher(s): Derek Friesen, Graduate Student, Educational Psychology and Special Education, University of Saskatchewan, Derek.W.Friesen@usask.ca

Researcher(s): Dr. Amin Mousavi, Assistant Professor, Educational Psychology and Special Education, University of Saskatchewan, Amin.Mousavi@usask.ca, 306-966-7653

Principal Investigator/Supervisor: Dr. Amin Mousavi, Assistant Professor, Educational Psychology and Special Education, University of Saskatchewan, Amin.Mousavi@usask.ca, 306-966-7653

Purpose and Objective of the Research:

We are doing a study to learn about how the concept of assessment literacy (AL) pertains to contexts in higher education (HE) settings. Our main purpose is to distinguish features of HE

settings necessitating an AL more sophisticated than can be described with current AL models, given how such models have largely been developed through investigation of teachers in school-based settings. In phase 1, an empirical scoping review of studies investigating assessment literacy resulted in a dataset of 182 research sources relating to knowledge, skills, and conceptions of assessment necessary for effective assessment in HE settings. We identified a small but notable body of literature related to student learning assessment literacy, but a paucity of literature related to non-student assessment literacy tasks in higher education. Themes emerging from these bodies of literature are analyzed in comparison to existing AL conceptualization and discussed considering prior AL research. In phase 2, we aim to develop an initial questionnaire to assess the AL of faculty members in Canadian HE institutions.

In order to complete this phase, the initial item pool will be sent to experts in AL with the goal of assessing its content validity. This an important step towards defining a unique framework of AL in HE which will allow for further enhancing the assessment practices in HE and providing a structure for targeted professional development for new and current faculty members in Canadian universities.

Procedures:

If you decide to take part in this study, you will participate by completing an item content review form. We will provide a review form containing our initial item pool via email, which you can also return via email. The form provides instructions for rating each item 1, 2, or 3, and provides space for feedback about item modifications. Participants can complete the review form in a location convenient for them. Completing the item review form will require an estimated 30-40 minutes. Please feel free to ask any questions regarding the procedures and goals of the study or your role.

Funded by:

This project was funded by a SSHRC grant.

Potential Risks:

There are no known or anticipated risks to you by participating in this research.

Potential Benefits:

Participant may derive personal satisfaction from contributing to research about assessment literacy in higher education settings.

Confidentiality:

Collected data will be used for survey development. Collected data will not be published. The survey, development process, and results will be reported on in scientific meetings, journal articles, and graduate theses. The information collected from you will be aggregated with other participants in all reports. No identifying information will ever be revealed in reporting of results. Moreover, the consent forms will be stored separately from the data so that it will not be possible to associate a name with any given set of responses.

Storage of Data:

Item review data will be securely stored by the graduate research assistant. No physical copies of the forms will be created. Electronic data will be stored using a password-protected cloud-based storage product (i.e., OneDrive) during analysis. Data will be stored for 5 years post-publication as per University of Saskatchewan guidelines. All data will be deleted by the PI after project completion and the minimum storage period.

Right to Withdraw:

You can participate or not. You can stop at any time. Your participation is voluntary, and you can provide as much feedback as you are comfortable with. You may withdraw from the research project for any reason, at any time without explanation or penalty of any sort. Should you wish to withdraw, you can contact the research team via email. Data is expected to be pooled by January 30th, 2021, after which point data withdrawal will no longer be possible.

Follow up:

To obtain results from the study, please contact the principal investigator. Results will be published in a graduate thesis and submitted to relevant educational assessment and measurement journals for peer review. If you wish to be notified directly about published results, please notify the researcher in your response.

Questions or Concerns:

Contact the researcher(s) using the information at the top of page 1. This research project has been approved on ethical grounds by the University of Saskatchewan Behavioural Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that committee through the Research Ethics Office: ethics.office@usask.ca; 306-966-2975; out of town participants may call toll free 1-888-966-2975.

Signed Consent:

Your signature below indicates that you have read and understand the description provided. I have had an opportunity to ask questions and my questions have been answered. I consent to participate in the research project. A copy of this consent form has been given to me for my records.

Name of Participant

Signature

Date

Researcher's Signature

Date

A digital copy of this consent will be retained by the research team. Please retain a copy for your own documentation.

Participant Consent Form

You are invited to participate in a research study entitled: Towards a Comprehensive View of Assessment Literacy for Higher Education Settings

Student Researcher(s): Derek Friesen, Graduate Student, Educational Psychology and Special Education, University of Saskatchewan, Derek.W.Friesen@usask.ca

Researcher(s): Dr. Amin Mousavi, Assistant Professor, Educational Psychology and Special Education, University of Saskatchewan, Amin.Mousavi@usask.ca, 306-966-7653

Principal Investigator/Supervisor: Dr. Amin Mousavi, Assistant Professor, Educational Psychology and Special Education, University of Saskatchewan, Amin.Mousavi@usask.ca, 306-966-7653

Purpose and Objective of the Research:

We are doing a study to learn about how the concept of assessment literacy (AL) pertains to contexts in higher education (HE) settings. Our main purpose is to distinguish features of HE settings necessitating an AL more sophisticated than can be described with current AL models, given how such models have largely been developed through investigation of teachers in school-based settings. In phase 1, an empirical scoping review of studies investigating assessment literacy resulted in a dataset of 182 research sources relating to knowledge, skills, and conceptions of assessment necessary for effective assessment in HE settings. We identified a small but notable body of literature related to student learning assessment literacy, but a paucity of literature related to non-student assessment literacy tasks in higher education. Themes emerging from these bodies of literature are analyzed in comparison to existing AL conceptualization and discussed considering prior AL research. In phase 2, we developed an initial questionnaire to assess the AL of faculty members in Canadian HE institutions using existing assessment literature and expert review.

In order to complete this phase, an electronic survey will be sent to faculty members in several Canadian HE institutes to provide a baseline data for faculty understanding, perception of, and involvement with tasks related to assessment literacy higher education in Canada. This an important step towards defining a unique framework of AL in HE which will allow for further enhancement of assessment practices in HE and provide a structure for targeted professional development for new and current faculty members in Canadian universities.

Procedures:

If you decide to take part in this study, you will participate by completing a survey about assessment literacy in higher education. We will provide a link to the online survey platform SurveyMonkey. The survey consists of [~40, dependent on expert review] items. Participants will respond to items using a Likert-type scale (i.e., 1 – 5) and a dichotomous scale (i.e., true or false). There is also an open-ended question. Participants can complete the survey using their

own computer. Completing the survey will require an estimated 20-30 minutes. Please feel free to ask any questions regarding the procedures and goals of the study or your role.

Funded by:

This project was funded by a SSHRC grant.

Potential Risks:

There are no known or anticipated risks to you by participating in this research.

Potential Benefits:

Participant may derive personal satisfaction from contributing to research about assessment literacy in higher education settings.

Confidentiality:

Collected data will be used for statistical analysis. Results of the survey will be reported on in scientific meetings, journal articles, and graduate theses. The information collected from you will be aggregated with other participants in all reports. No identifying information will ever be revealed in reporting of results. Moreover, the consent forms will be stored separately from the data so that it will not be possible to associate a name with any given set of responses. This survey is hosted by Survey Monkey. Your data will be stored in facilities hosted in Canada. Please see the following for more information on [Survey Monkey's Privacy Policy](#).

Storage of Data:

Survey data will be stored using Survey Monkey. Additionally, a copy of the raw data will securely stored by the principal investigator. During analysis, electronic data will be stored using a password-protected cloud-based storage product (i.e., OneDrive). After analysis, data will be stored for 5 years post-publication as per University of Saskatchewan guidelines. All data will be deleted by the PI after project completion and the minimum storage period.

Right to Withdraw:

You can participate or not. You can stop at any time. Your participation is voluntary, and you can stop the survey at any point until submitting the completed survey. Because the survey is anonymous, data withdrawal will not be possible after you submit the completed survey. Data collection will resolve by February 2021, after which data analysis will commence.

Follow up:

To obtain results from the study, please contact the principal investigator. Results will be published in a graduate thesis as well as submitted to relevant educational assessment and measurement journals for peer review. If you wish to be notified directly about published results, please notify the researcher in your response. It is expected that results will be available by June 2021.

Questions or Concerns:

Contact the researcher(s) using the information at the top of page 1. This research project has been approved on ethical grounds by the University of Saskatchewan Behavioural Research Ethics Board. Any questions regarding your rights as a participant may be addressed to that

committee through the Research Ethics Office: ethics.office@usask.ca; 306-966-2975; out of town participants may call toll free 1-888-966-2975.

By completing and submitting this questionnaire, **your free and informed consent is implied** and indicates that you understand the above conditions of participation in this study.

Appendix H: Initial Item Pool

Necessary Knowledge and Skills

Proposed Prompt: How knowledgeable do people in your faculty/department/discipline need to be about each aspect of assessment below?

Proposed response scale: 5-point Likert scale:

Not knowledgeable at all/slightly knowledgeable/moderately knowledgeable/very knowledgeable/extremely knowledgeable

Item
Accommodating assesses with disabilities or other impairments
Aligning tests to proficiency frameworks
Designing scoring keys and rating scales (e.g., rubrics) for assessment tasks
Determining pass-fail marks/cut scores
Developing specifications (overall plans) for assessments
How to train others about assessment
Identifying assessment bias
Making decisions about what to assess
Piloting/trying out assessments before their administration
Selecting appropriate items or tasks for a particular assessment purpose
Selecting appropriate rating scales
Training others to use rating scales (e.g., rubrics) appropriately
Training others to write good quality items (questions) or tasks for assessments
Writing good quality items (tasks) for assessments
How to give useful feedback on the basis of an assessment
How to use assessments to diagnose learners' strengths and weaknesses
How to use assessments to guide teaching and learning goals
How to use assessments to motivate student learning
How to use peer-assessment
How to use self-assessment
How assessment is used in my department*
How assessments can be used to enforce social policies
How to determine if an assessment aligns with a local educational system
How to determine if an assessment aligns with a local system of accreditation
How to determine if the results from an assessment are relevant to the local context
How to identify stakeholders in local assessment contexts
How to use assessment in everyday tasks
How to use assessment strategies that are typical for my discipline
Institutional goals and guidelines about what to assess
The assessment traditions in your local context
The relevant legal regulations for assessment in your local area
How one's beliefs or attitudes are shaped by life circumstances such as cultural background
How one's beliefs or attitudes may conflict with those of other groups involved in assessment
How one's interpretations could be different depending on who/what is being assessed
How one's own beliefs or attitudes might influence one's assessment practices
How one's own knowledge of assessment might be further developed
One's own beliefs or attitudes towards assessment

Using statistics to analyze overall scores on a particular assessment
Using statistics to analyze the difficulty of individual items (questions) or tasks
Using statistics to analyze the quality of individual items/tasks
Using techniques other than statistics (e.g., questionnaires, interviews, analysis of language) to get information about the quality of an assessment
How to interpret what a particular score means in context
How to use assessments to evaluate achievement in learning
The concept of reliability (how accurate or consistent an assessment is)
The concept of validity (how well an assessment measure what it claims to measure)
How assessments can influence teaching and learning in the classroom
How assessments can influence teaching and learning materials
How assessments can influence the design of a course or curriculum
How to prepare learners to take assessments
Scoring closed response questions (e.g., Multiple Choice Questions)
Scoring open-ended questions (e.g., short answer questions)
Using rating scales to score assessments
How one's assessment knowledge shapes their opinion of assessment
How one's opinion of assessment shapes their knowledge of assessment
What theory or concepts one does not know about assessment
What theory or concepts one knows about assessment
How to clarify the meaning of assessment concepts used when discussing assessment
How to communicate about assessment results and decisions
How to speak about assessment with colleagues in department
How to speak about assessment with stakeholders outside of department
How one's assessment experiences affect their assessment practices
How to find opportunities to develop my assessment literacy
How to keep up to date with assessment research
How to participate in community learning activities
How to reflect on one's assessment knowledge, practices, and experiences
Where to access relevant literature about assessment
How to use assessment in one's professional tasks
How to use assessment in professional decision-making
How to incorporate assessment in one's day-to-day work
How to evaluate professional decisions

Item
Assessment of student's learning
Assigning course grades
Assigning final grades
Developing student assessments
Evaluating learning materials such as textbooks or research artifacts
Evaluating student assessment instruments
Instrument assessment: critique validity and reliability of tools
Scoring classroom assessments

Student admissions decisions
Student feedback
Student graduation decisions
Student learning outcomes assessment
Student program selection decisions
Establishing baseline program criteria
Leading departmental assessment efforts
Program meta-assessment: assessing assessment processes
Program needs assessment
Program Outcomes Assessment
Program review
Using assessments to evaluate programs
Accountability/external evaluation initiatives
Advising (students, projects, etc.)
Collaborating outside of department
Collaborating within department
Communicating assessment outcomes
Evaluation of faculty research products/artifacts
Evaluation of theses/dissertations
Faculty evaluation using self-assessment
Faculty evaluations based on peer assessment
Faculty evaluations based on student evaluations
Faculty hiring decisions
Faculty promotion decisions
Faculty tenure decisions
Leading/coordinating departmental assessment efforts
Participating in institutional assessment initiatives
Serving on committees
Conducting research
Disseminating research
Literature review/ assessing scholarly research
Managing complex projects
Scale development
Calculating measurement error
Interpreting measurement error
Calculating a score
Interpreting what a particular score means
Assessing test item's quality (e.g. item difficulty, item discrimination)
Language assessments
Determine if assessment aligns with local system of accreditation
Please list any additional assessment responsibilities:

Perceived roles for assessment in HE

Proposed Prompt: Please indicate if you believe the following statements to be true or false.

Proposed response scale: Dichotomous scale: True/false

Item
Assessment processes result in information that is useful to me
For the most part, assessment processes are the same across contexts
I have a stable theoretical understanding of assessment
I have the theoretical knowledge necessary to perform assessment-related tasks in my position
The same assessment principles underlie any assessment task
Depending on who I am assessing, my interpretations could be different
I am able to put my assessment knowledge into practice for any given task
I have a pragmatic working knowledge of assessment theory
I recognize professional situations where I can apply my assessment knowledge
Similar assessment concepts can be applied to different situations
I am aware of what I know about assessment
I believe assessment is common sense.
I reflect on my assessment knowledge and practices
My opinion of assessment is shaped by my knowledge of assessment
To do my job, I need to know how my knowledge shapes my beliefs
Assessment is more of an obligation
I belief assessment is valuable
I don't need to use assessment for certain people because the outcome is obvious
I regard assessment as useful in my profession
My assessment efforts are generally worthwhile
My views of assessment have been shaped by experiences
I clarify the meaning of the assessment concepts I use when discussing assessment with others
I get confused when discussing assessment with my colleagues
I know how to communicate about assessment results and decisions
I understand the concepts other people use when discussing assessment
I use the term assessment in the same way as colleagues from other departments
I use the term assessment in the same way as colleagues in my department
I can determine if the results from an assessment are relevant to the local context
I can identify stakeholders in local assessment efforts
I completely follow assessment guidelines set by my institution
I understand how assessment is used in my department/college/faculty
I use assessment strategies that are typical for my professional discipline
My department/college/faculty shapes what and how I assess
My institution sets guidelines about what and how to assess
Assessment promotes consistency in decision-making
I use assessment when I make professional decisions
In higher education, decision making is not possible without assessment skills
I do not need assessment skills to perform my tasks
I need to keep up to date with assessment research to perform professional tasks
I participate in community learning activities regarding assessment
I seek out opportunities to increase my assessment literacy
My understanding of assessment is up to date with research literature
I am confident at using student assessment, but not other kinds of assessment

I am engaged with assessment processes in my profession
I am experienced at using assessment
I consider myself to be an assessor
I do not feel confident using assessment
I do not need to practice assessment for making decisions
I feel ownership over when and how to use assessment