Norming at Scale:

Faculty Perceptions of Assessment Culture and Student Learning Outcomes Assessment

by

Nicholas Williams

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

Approved March 2018 by the
Graduate Supervisory Committee:

Daniel Dinn-You Liou, Chair
Erin Rotheram-Fuller
David Turbow

ARIZONA STATE UNIVERSITY

May 2018

ProQuest Number: 10809240

ProQuest

ProQuest 10809240

ABSTRACT

To foster both external and internal accountability, universities seek more effective models for student learning outcomes assessment (SLOA). Meaningful and authentic measurement of program-level student learning outcomes requires engagement with an institution's faculty members, especially to gather student performance assessment data using common scoring instruments, or rubrics, across a university's many colleges and programs. Too often, however, institutions rely on faculty engagement for SLOA initiatives like this without providing necessary support, communication, and training. The resulting data may lack sufficient reliability and reflect deficiencies in an institution's culture of assessment.

This mixed methods action research study gauged how well one form of SLOA training – a rubric-norming workshop – could affect both inter-rater reliability for faculty scorers and faculty perceptions of SLOA while exploring the nature of faculty collaboration toward a shared understanding of student learning outcomes. The study participants, ten part-time faculty members at the institution, each held primary careers in the health care industry, apart from their secondary role teaching university courses. Accordingly, each contributed expertise and experience to the rubric-norming discussions, surveys of assessment-related perceptions, and individual scoring of student performance with a common rubric. Drawing on sociocultural learning principles and the specific lens of activity theory, influences on faculty SLOA were arranged and analyzed within the heuristic framework of an activity system to discern effects of collaboration and perceptions toward SLOA on consistent rubric-scoring by faculty participants.

Findings suggest participation in the study did not correlate to increased inter-rater reliability for faculty scorers when using the common rubric. Constraints found within assessment tools and unclear institutional leadership prevented more reliable use of common rubrics. Instead, faculty participants resorted to individual assessment approaches to meaningfully guide students to classroom achievement and preparation for careers in the health care field. Despite this, faculty participants valued SLOA, collaborated readily with colleagues for shared assessment goals, and worked hard to teach and assess students meaningfully.

DEDICATION

To Vito and Wally:

Daddy loves you so much; like the song says, "All that I can give you is forever yours to

keep." And I can't wait for all the fun we'll have, but for now, let's all promise to get a

little more sleep.

ACKNOWLEDGMENTS

entering my doctoral program, he made me feel accepted as an assessment professional and inspired me with his expertise. Dr. Erin Rotheram-Fuller always greeted my work with enthusiasm. She has a knack for delivering calm, logical feedback, which often resulted in sorely-needed boosts of confidence for me. I am especially grateful for her help in shaping stronger, clearer design elements of my study. Finally, my chairperson, Dr. Daniel Dinn-You Liou, is owed tremendous gratitude for his consistent and caring mentorship. Throughout my research, Dr. Liou challenged me to work harder, think more critically, and explain more clearly. He is a brilliant example of someone driven to succeed, and I benefited tremendously from his model work ethic, writing quality, and clarity of thought. Thank you for pushing me to such heights.

Finally, I must acknowledge my family for their unconditional encouragement, love, and support. I've lived in Phoenix many years now, but any bit of success I'll ever enjoy can be traced back to the strong Illinois roots that have shaped me. Maggie, Frank, Mom, and Dad - thanks for putting up with me for so long. I love you, and I'm proud to be your brother and son. And to my wife, Robyn – WE did it. I am so blessed to share this journey with you; thanks for the amazing sacrifices you made to help me accomplish this. I can't wait for all the laughter and beauty we will continue to create together, and I love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

In 2006, the US Department of Education commissioned a report on the state of U.S. education, including higher education. In the resulting work, now referred to as the Spellings Commission Report, experts strenuously recommended schools enhance student learning outcomes assessment (hereafter, SLOA), and they urged accreditors to explicitly seek evidence of effectiveness of SLOA measures from their member institutions (U.S. Department of Education, 2006). Within higher education, assessment may refer to many different uses enacted over several decades (Astin, 1991). According to Banta & Palomba (2014), outcomes assessment means to evaluate curricular or institutional effectiveness, but SLOA is more specifically defined as "the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving student learning and development" (Banta & Palomba, 2014, p.1-2). The Spellings Commission's findings encouraged a vein of emerging scholarship to assist educators in defining meaningful student learning and methods for learning assessment (Darling-Hammond, Wilhoit, & Pettinger, 2014; Kezar 2013). Since then, authors have focused on key components of SLOA in higher education: insisting faculty be central in SLOA work, that multiple stakeholders collaborate on assessment projects, and that SLOA be explicitly linked to and supported by an institution's mission (Banta & Palomba, 2014; Kuh et al., 2015; Suskie, 2009). These elements, which each speak to the co-constructed nature of learning expectations, and of SLOA more specifically, will each be explicated further in this dissertation.

Also in line with the Spellings Commission's recommendations, other researchers studied the impact of enhancing validity and reliability to promote effective SLOA (Jonsson & Svingby, 2007; Royal, 2011). Reliability refers to the consistency with which assessment measurements are collected, and validity is the extent to which these measurements accurately capture the constructs intended (Waugh & Gronlund, 2013). Along with increased regulatory focus and scholarly research, regional and program accrediting bodies have disseminated clear expectations for institutions demonstrating reliability and validity of assessment methods (Eldridge, 2016; Higher Learning Commission, 2012; Southern Association of Colleges and Schools Commission on Colleges, 2012). Professional associations within higher education have supported these accreditation mandates, with numerous publications describing the importance of consistency and overall quality in student learning outcomes assessment data (American Educational Research Association et al., 2014; Kuh, Jankowski, Ikenberry, & Kinzie, 2014; Miller & Leskes, 2005).

These dovetailing efforts toward increased accountability also gave rise to the term *culture of assessment*. One set of authors defines it as the "institutional contexts supporting or hindering the integration of professional wisdom with the best available assessment data to support improved student outcomes or decision making" (Fuller, Skidmore, Bustamonte, & Holzweiss, 2016, p.404). In other words, every institution has a culture of assessment, positive or negative, that describes the extent to which a school practices effective, inclusive SLOA and bases its efforts in continuous improvement of student learning rather than compliance. A college's assessment culture can be captured by examining stakeholder perceptions of various assessment-related structures at an

2

institution (Fuller et al., 2016; Guetterman & Mitchell, 2016; Holzweiss, Bustamonte, & Fuller, 2016; Ndoye & Parker, 2010). In this study, perceptions of assessment will be studied particularly through the lens of faculty due to their central role in affecting student learning.

Though garnering more focused empirical research in recent years, the term *culture of assessment* has long lacked a comprehensive definition, all too often connoting formulaic revamps of systems and processes within an institution for accreditation or regulatory needs (Kezar, 2013). These unfortunate forms of external accountability constrain colleges with unwieldy, overly prescriptive methodologies (McClellan, 2016). Indeed, professional and accreditor standards for SLOA constitute institutional isomorphism: generic guidance leading some schools to adopt ill-fitting models in order to comply with procedural assumptions or prestigious ideals for higher education (Farquharson, 2013; Toma, 2008). This is especially true in the private, for-profit sector, where colleges and universities rely on accreditation credentials and other marks of quality from external bodies to establish the credibility of their programs (Kinser, 2005; 2007). Much of the standardized procedures for accountability discussed here developed in response to recent shifts in the average student, faculty, and institutional profiles have changed – a more diverse population of people engage in higher education seeking more specifically tailored experiences (U.S. Department of Education, 2015). Along with these sociocultural phenomena, some experts argue SLOA and our common understanding of learning should shift too (Shulman, 2007). In doing so, an institution's culture of assessment evolves with changing accountability structures.

Another component of such an assessment culture meriting mention here is collaboration among educators. Because of the aforementioned diversity found in today's higher education arena, the ability to work productively with others retains high value. Collaboration as a research phenomenon can be found in a wide range of educational contexts: studies of classroom teachers co-constructing instructional goals (Goodnough, 2016; Pietarinen, Pyhältö, & Soini, 2016), higher education faculty working together to build long-term engagement and accountability (Haviland, Shin, & Turley, 2010) and administrators at all levels investing in social capital of their educators (Hargreaves & Fullan, 2012). These studies symbolize the idea that collaboration means much more than simple interaction among educators. In this study collaboration will be viewed through the lens of relational agency, a sociocultural concept characterized as the ability to view others' perspectives as useful, to allow one's own views to be influenced by others', and the ability to work with others to craft common goals and products (Edwards, 2005).

Scholars of effective student learning outcomes assessment or culture of assessment identify collaboration and inclusivity as centrally important to sustained success (Banta & Palomba, 2014; Suskie, 2009). Nearly any accountability measure enacted by a university or college seeks catalysis from collective action and a shared vision of those involved (Maki, 2010). Collaboration drives positive accountability efforts in education by inviting all stakeholders to co-construct professional learning standards (Fuller, 2012; O'Connell et al., 2016). As a shared exercise then, inclusive, substantial dialogue with educators helps a college or university ask the right questions about what its student learn (Astin, 1991; Cordero de Noriega & Diiorio, 2006). Within a

4

classroom or among a small group of professionals, this critically reflective work allows researcher-educators to narrow in on shared goals (Kogan, Conforti, Bernabeo, Iobst, & Holmboe, 2015; Guetterman & Mitchell, 2016). At scale and over time, critical reflection generates action-oriented cycles of inquiry to gauge whether interventions achieved their aims (Herr & Anderson, 2015; Schoepp & Benson, 2016). In the case of this study, collaboration and faculty perceptions of SLOA are explored as influences on rubric-norming workshops seeking to imbue consistency – and teach the importance of consistency – in the SLOA done by faculty who share common scoring instruments.

Pressing Problem of Practice

In response to demands from our regional accreditor, university leaders tasked my Institutional Effectiveness team with overseeing an institution-wide SLOA data collection pilot in late 2013 and 2014. We successfully laid the groundwork for a comprehensive assessment system driven by faculty-scored rubrics in the classroom, but the effort was rushed, motivated chiefly by external accountability. Of greatest concern were the hundreds of faculty members who received no training to interpret student work reliably when using common rubrics. This resulted in a lack of consistency that threatened the validity of inferences drawn from SLOA data summaries. Good measurement practice assumes all faculty raters apply a common scoring rubric uniformly when judging the same student performance (Judd, Secolsky, & Allen, 2012; Maki, 2010). Moreover, this inconsistency of SLOA data quality and lack of assessment training emblematizes overall areas for growth within the university's assessment culture, especially regarding faculty perceptions of SLOA. Research shows training can help higher education faculty adopt a better understanding of SLOA's role in improving

student learning (Danley-Scott & Scott, 2014; Kogan et al., 2015) and also reinforce a sense of belonging with the institution's mission and vision (Haviland, 2014). Effective assessment training, focused on raters' use of common instruments, can also reduce variance in scoring interpretations (Jonsson & Svingby, 2007; Rezaei & Lovorn, 2010) especially when such training emphasizes professional skill development, collaboration, and critical reflection (Kogan et al., 2015). The training interventions employed in this study fostered these critical elements of a healthy assessment culture with a series of interactive faculty workshops.

Local Context

Southwest University, or SWU, symbolizes many of the demographic and operational shifts seen throughout collegiate academics these days. Though unique in many ways, SWU shares with other schools the burden of navigating a changing, often uncertain higher education landscape. Because of this, SWU offers a critical lens through which to view the overarching challenge facing colleges, to conduct authentic, rigorous, and externally compliant learning assessment of students in a consistent manner across an institution.

School Profile. A private, for-profit institution, Southwest University conducts over 80% of its courses online but maintains active brick-and-mortar campuses in several US states. Administrative units in business intelligence, academic operations, and institutional effectiveness oversee logistics for the university. Other staff units assist faculty and students with technological support, compensation/finances, and advising. Academically, SWU offers more than a hundred certificate and degree programs at the associate, bachelor's, master's, and doctoral levels, spanning a wide variety of subject

areas. Its most popular programs reside in the subject areas of business, information systems and technology, and health professions, but the university aspires to career-relevance for its entire academic portfolio. In support of this, the university and each of its colleges maintain a leadership hierarchy to craft, publish, and support standardized curriculum for all SWU campuses. Historically less common in higher-learning institutions, this set-up for curricular design has become increasingly necessary at multi-site institutions like SWU given the scale of the university's campus network (Kinser, 2005; 2006).

The massive organizational structure at SWU supports a student base numbering over 100,000. SWU's student profile bears little resemblance to common notions of 18-to-22 year-old students living on campus (U.S. Department of Education, 2015). SWU students, on average, are over the age of 35, claim heritage in a racial or ethnic minority, work full-time, are the first in their family to attend college, and support a household with at least one dependent. All SWU students are considered "full-time" (U.S. Department of Education, 2014), taking single, intensive courses of five to eight weeks in duration in succession.

To serve its enormous, diverse student population, the university employs thousands of faculty members spread throughout the United States. Nearly all SWU faculty work for the university part-time, holding full-time jobs in the same professional areas that they teach. The majority teach exclusively online, but all utilize an online learning management system for instruction at the for-profit institution. Beyond those uniform characteristics, however, lay a general unevenness: a wide spectrum of subject-matter knowledge, schooling background, teaching efficacy, and overall engagement with

7

university culture. Internal SWU faculty surveys reveal scores of contented, productive faculty next to many disenfranchised colleagues who feel underpaid, underappreciated, and ineffectual within the larger organization. These findings match other research across the higher education landscape. Part-time and adjunct faculty positions, as well as online teaching environments, continue to increase across higher education (Selingo, 2016). Part-time faculty indicate low levels of perceived institutional support (U.S. Department of Education, 2002), demonstrate less instructional efficacy than their full-time counterparts (Schuetz, 2002). Other studies indicate faculty performance at for-profit institutions may be hindered by a lack of academic freedom and the conflictual nature of their institution's profit motives with teaching and learning (Lechuga 2008; 2010). Regardless of an institution's structure, though, all schools rely on their faculty for assistance in gathering evidence of student learning outcomes or for another accountability exercise (Kezar & Maxey, 2014). Although one cannot generalize to an entire faculty base at a higher education institution, these insights lead toward a greater understanding of educators' work with SLOA and their perceptions of SLOA.

Assessment Culture at SWU. The Office of Institutional Effectiveness (OIE) helps enact accountability measures for all academic and co-curricular programs within the university. This includes oversight for data collection, analysis, and reporting for all SLOA measures in coordination with the schools and colleges. This work includes a standardized, assessment-focused process for creating mission statements, curriculum maps, and student learning outcomes for each academic program. Within the university's centralized curriculum-design process, assessment staff assist SWU schools and colleges with developing specific mastery-level assignments throughout program sequences that

aligned to learning outcomes. The process allows faculty to apply a rubric to student work in the classroom and submit scores to a central assessment management tool. Once collected and aggregated, the scores theoretically measure the overall level of student achievement toward learning outcomes.

Common use of rubrics by faculty, however, must be supported by instrument-specific training to imbue sufficient inter-rater reliability into the measurements (Jonsson & Svingby, 2007). Faculty training did not include this content initially, and all preceding systemic change resulted from top-down, compliance-focused mandates and limited faculty involvement. Experts state definitively that faculty should possess a central role in assessing student learning outcomes (Banta & Palomba, 2014; Fuller et al., 2016), and that internal accountability initiatives drive real, sustainable change (Fullan, Rincón-Gallardo, & Hargreaves, 2015). When these conditions did not readily appear in the assessment changes at SWU, it accurately reflected areas for growth within the institution's culture of assessment. From our team's perspective, we experienced reluctant engagement and lower overall buy-in from faculty, coherent with recent scholarship documenting faculty wariness and skepticism when SLOA initiatives were not accompanied by proper support or explanation (Cain & Hutchings, 2015; Danley-Scott & Scott, 2014). At a large institution like SWU, with a multitude of experiences and assessment interpretations, it was hypothesized that a 'norming' intervention, scaled out to all colleges and programs, could work to harness diverse faculty perspectives toward a more consistent set of student learning expectations.

9

Role of the Researcher

In late 2014, I asked to extend our centralized faculty assessment training to include rubric-specific content for faculty raters using common scoring instruments. The request, based on a preliminary literature review (Jonsson & Svingby, 2007) and my own education measurement studies, was received well by leadership. Soon after, we had created a system by which faculty could be compensated for two or three hours of rubric-specific training with colleagues. With SWU schools and colleges, I helped design the content and train the faculty liaisons to lead sessions. At the same time, I grew interested in a conducting a formal research study while developing dual positionality as both insider and outsider to the situation.

When an action researcher examines the complexity of this dual positionality, numerous potential dilemmas merit consideration along with any benefits (Herr & Anderson, 2015). My active involvement with assessment training and systems has afforded me specific content knowledge and familiarity with like-minded personnel. This 'insider' perspective as an assessment professional helps me discover opportunities for collaboration and support. One less familiar with higher education assessment, and assessment at SWU, would not enjoy the same resources. I had to ensure these privileges did not obscure a pragmatic interpretation of our institution's nascent assessment culture during this research. On the other hand, my study participants likely self-identified as faculty members within a single academic program and thus viewed me as 'outsider'. In other training situations, I emphasized my knowledge or skills in the training material because I felt inferior to faculty members for lacking their college classroom experience. When instructors trade stories or allude to particular pedagogical issues in workshops, I

10

do not share their frame of reference. To overcome this feeling, I have attempted to mitigate my non-belonging by providing transparent communication about the need for SLOA and promoting a shared sense of engagement in SLOA training. I expected that, in doing so, I would help foster aspects of a healthier culture of assessment (Fuller et al., 2016; Guetterman & Mitchell, 2016), and I hoped to open myself to a clearer understanding of faculty perceptions of SLOA.

Purpose Statement and Research Questions

The purpose of this action research project was to examine how rubric-specific 'norming' workshops for faculty raters impact specific aspects of the institution's culture of assessment: consistency of SLOA with rubrics, faculty collaboration in SLOA-related rubric-norming sessions, and faculty perceptions of SLOA. The following research questions guided the study. The first and second questions explored the intervention's effect on inter-rater reliability and perceptions of assessment within an institution. The third question probed faculty raters' ability to craft shared expectations for student learning through collaboration in the training environment. The fourth question sought to understand how 'norming' collaboration and perceptions of SLOA mediate faculty's ability to consistently and effectively assess student learning outcomes.

1) To what extent does rubric-rater training improve inter-rater reliability among faculty scoring student performance assessments in a health administration course?

2) To what extent does rubric-rater training strengthen faculty perceptions of student learning outcomes assessment?

3) How do faculty collaborate with one another when assessing student benchmark assignment work in a health administration course?

4) How do faculty collaboration and perceptions of assessment mediate their consistent assessment of student learning outcomes?

CHAPTER 2

THEORETICAL FRAMEWORKS AND RESEARCH GUIDING THE PROJECT

This study sought to understand how faculty assessment of student learning outcomes, and consistency of assessment, is impacted by faculty perceptions of student learning outcomes assessment (SLOA) and assessment-related collaboration within an institution's culture of assessment. Literature on SLOA and culture of assessment is reviewed, showing how both constructs intertwine as a university measures student learning. This is followed by an explication of rubric use for SLOA in higher education, and how rubrics function as a tool to enhance assessment data quality. Rubric norming is next located in the literature, as a research-based intervention method for the study. Key elements of norming as a training intervention are identified. Then, two major factors on norming efficacy are discussed: faculty perceptions toward SLOA and collaboration among educators, especially for assessment purposes. Faculty perceptions toward SLOA and training toward effective assessment of student learning outcomes with rubrics are grounded in other recent studies. Relational agency, a construct that illuminates the capacity of faculty to view their work as collective and interpret the perspectives of others when collaborating through SLOA training, is next introduced. As explained in the chapter, these two influences constitute the primary research focus for data collection and analysis.

Following the literature review of terms relevant to the study, the researcher's theoretical framework is introduced and explained. Activity theory provides a lens for viewing SLOA by faculty as a complex task mediated by the rubrics used at SWU for assessment work, faculty perceptions of assessment, and collaboration encouraged by the

13

rubric-norming intervention. Activity systems are described as a heuristic helping to understand complex and interdependent influences within a problems of practice, especially in educational contexts. In sum, activity theory establishes an effective framework for studying faculty perceptions of SLOA and faculty collaboration as two of many contextual factors impacting the effectiveness of rubric norming for faculty SLOA.

SLOA and Culture of Assessment

Assessment of student learning outcomes, a specific brand of educational measurement, has grown in importance in the last quarter-century of higher education (Banta & Palomba, 2014; Suskie, 2009). As defined in the first chapter, SLOA signifies a process of gathering quality information about an educational program and how it cultivates student learning and development (Fuller, 2012). The standardization of SLOA practices by regional and professional accreditors in recent decades has resulted in schools pursuing more rigorous assessment methods to gauge learning and effectiveness (Kezar, 2013). Such methods balance a focus on external accountability – being answerable to one's accreditors and other outside stakeholders – with an equal measure of internal accountability and professional pride (Bresciani, 2011; Fullan et al., 2015). To attain the proper balance, institutions engage an array of stakeholders to create a shared definition of SLOA and incorporate it into the mission, vision, and values of an institution (Banta & Palomba, 2014; Fuller et al., 2016). Leaders can then focus their educators on consistent and accurate data-gathering, to address areas of need with targeted improvements over iterative assessment action cycles. As SLOA becomes more effective, purposeful and inclusive, a school develops positive aspects of its 'culture of assessment' (Fuller et al., 2016; Ndoye & Parker, 2010; Suskie, 2009; Weiner, 2009).

14

Culture, not climate. Culture more generally has been defined, for an organization or professional system, as a shared set of behaviors, values, and norms learned and perpetuated by members of the group in question (Schein, 2010). Scholars of organizational culture describe 'layers' of culture moving in a hierarchy from artifacts of culture – the outward demonstration of shared belief – to shared behaviors to the deep, underlying norms and assumptions driving the culture (Rousseau, 2010; Schein, 2010). Recent scholarship on organizational culture has attempted to clarify differences between 'culture' and 'climate' from methodological and conceptual standpoints (Schneider, Ehrhart, & Macey, 2013). Education researchers have echoed these definitions. One recent study recognized school culture and school climate as explicitly different constructs, with climate encompassing a broader but more fleeting connotation of one's perceptions as related to environmental factors while culture remains more deeply embedded in one's beliefs (Van Houtte & Van Maele, 2011). In both cited studies, researchers acknowledge a general lack of agreement and a prevailing inclination by the research community to conflate the two. An argument can be made, however, that studying culture is ultimately more conducive for looking at school effectiveness (Van Houtte, 2005), and scholars see culture as an easier concept to measure empirically as related to other study phenomena (Schneider et al., 2013). In this study, 'culture' remained the focus, due to the aforementioned qualities of the construct and its predominant use in the professional literature related to SLOA.

Culture of assessment. Within higher education assessment, the construct of organizational culture has been adopted to evaluate an institution's commitment to making program decisions informed by evidence of student learning (Maki, 2010).

15

Lakos & Phipps (2004) explained the term 'culture of assessment' with organizational culture as the basis, advising academic librarians to build on existing administrative or structural strengths as a way to foster assessment-culture characteristics in current practice. More recent scholarship on 'culture of assessment' depicts it as an ideal state to be achieved or as an institutional re-emphasis of traits representing basic SLOA practice (Weiner, 2009). Other authors have empirically measured characteristics of assessment culture, surveying school leaders for their perspective (Fuller et al., 2016). All assessment experts writing about assessment culture generally agree on the fundamental concepts underpinning both quality SLOA and strong culture of assessment: a clear centrality for faculty and staff involvement, shared use of assessment data, common mission or values driving assessment work, and clear leadership for assessment efforts (Fuller et al., 2016; Fuller & Skidmore, 2014; Gorran Farkas, 2013; Ndoye & Parker, 2010). That multiple experts agree on a general set of assessment-culture factors substantiates their impact on SLOA, but it also creates a problem. One scholar notes the term has "become every aspect of the organization, making it both meaningful and meaningless" (Kezar, 2013, p.192). 'Culture of assessment' now connotes a soft buzzword to many, too ubiquitous and loosely defined to be well understood as a lever for effective SLOA and SLOA-based decision making.

The current study built on these works in two specific ways, studying assessment culture through the lens of faculty rather than school leaders, and by operationally defining the term clearly. By funneling one's understanding of SLOA through faculty perceptions, the study aligns with relevant research that definitively states faculty investment is essential to SLOA success (Cain & Hutchings, 2015). Guetterman and

Mitchell (2016) focused their assessment-culture inquiry on faculty perspectives, using a mixed-methods approach to study faculty-leader participants in a year-long professional development program specifically for learning outcomes assessment. Using surveys, qualitative feedback, and a summative poster project created by faculty participants in the program, the authors discovered faculty craved additional support and resources from their administrators and were most positive regarding the ability to learn collaboratively with – and from – their colleagues (Guetterman & Mitchell, 2016). Survey instruments helped the authors reinforce specific constructs which they described as constituting assessment culture. One specific survey gauged faculty perceptions toward SLOA in three parts: knowledge about assessment; personal dispositions toward SLOA; and perceived institutional encouragement for both SLOA and institutional use of assessment insights (Guetterman & Mitchell, 2016). This configuration of assessment-culture principles, as perceived by faculty, effectively synthesizes aforementioned relevant literature, which had narrowed on a similar set of composite characteristics to describe a culture of assessment: administrative leadership or support; faculty engagement and investment; use of SLOA data; sharing or communication norms; and shared purpose or vision for assessment at an institution. Condensed into the Guetterman & Mitchell model, these characteristics will be explored using multiple data collection methods described in Chapter 3, and represent a significant but limited portion of an institution's overall assessment culture. Now having defined 'culture of assessment' and planned its exploration through the lens of faculty training participants' perceptions, further explanation of SLOA is warranted.

Measuring Student Learning Outcomes with Rubrics

SLOA can be achieved using a variety of methods, but classroom work "assigned by faculty has always been the most meaningful and natural source of evidence for documenting student learning" (Hutchings, Kinzie, & Kuh, 2015, p.34). In higher education, most program-level student learning outcomes comprise complex, higher-order thinking skills (Banta & Palomba, 2014). Because of this, educators use performance assessments and rubrics to accurately assess those skills (Jonsson & Svingby, 2007; Lovorn & Rezaei, 2010; Reddy & Andrade, 2010). Rather than rely on a student's self-reported abilities, or abstractly gauging them with an objective test, performance assessments require students to demonstrate mastery of a learning outcome through direct and authentic application of the skill (Hutchings, et al., 2015). A rubric allows an educator to deconstruct a mastery performance into a set of dimensions or criteria, with each component organized along a continuum of ability. Faculty then rate the learning for each criterion of the demonstration, using their expertise and professional experience (Suskie, 2009; Walvoord, 2004). Instead of binary right-or-wrong outcomes as with traditional examinations, rubrics create multi-faceted student learning data. Educators have increasingly relied up on these scoring instruments, hoping to distill rich learning experiences into informative but digestible summaries (Brookhart & Chen, 2015).

In response to the emerging trend of rubric use in higher education, researchers have concentrated much scholarship at understanding the core components of rubrics (Dawson, 2015; Hack, 2015) and how educators may enhance student learning with them (Jonsson, 2014). Other researchers have conducted meta-analyses of the resulting

18

literature, in an attempt to summarize the growing number of studies focused on rubrics as tools to capture student learning in higher education (Brookhart & Chen, 2015; Jonsson & Svingby, 2007; Reddy & Andrade, 2010). Jonsson and Svingby (2007) emblematize the general theme of reviews that succeeded it. By reviewing 75 different published studies employing rubrics, the researchers examined if other researchers provided evidence of validity and reliability for rubric use, and if several hypothesized benefits of rubrics held substantial footing in the literature. Jonsson and Svingby conclude that future rubric research should focus on data quality, so that reliability and validity claims can be fully supported (2007). Further, they assert that rubrics can facilitate efficacious SLOA, beneficial to both students and faculty, but training and proper design principles should be followed (Jonsson & Svingby, 2007). The literature reflects a general belief that, with effective training and communication, rubrics can help more accurately gauge student learning at an institution (Jonsson & Svingby, 2007).

Consistency through rubric training. Perhaps the most common principle of rubric use found in literature is that, when a scoring instrument will be utilized commonly by multiple faculty raters, those raters must be trained rigorously (Lovorn & Rezaei, 2010; Sadler, 2005; Saxton, Belanger, & Becker, 2012) and some measure of consistency, or inter-rater reliability, must be sought (Kuh et al., 2015; Oakleaf, 2009; Royal, 2011; Turbow & Evener, 2016). Reliability estimates help validate inferences resulting from data collection (Judd et al., 2012; Stemler, 2004). Saxton et al. (2012) found that, especially for complex constructs such as critical thinking, meticulous attention to rater bias and other details was imperative for success. Their study tracked two graduate-student raters scoring three hundred student rubric scores using a

19

researcher-designed assignment and previously-conceived critical thinking rubric. The authors explain how valid and reliable scoring instruments, when implemented through a well-conceived plan for training and preparing raters, can achieve acceptable levels of inter- and intra-rater reliability (Saxton et al., 2012), albeit for only the pair of raters. Also transferable from the study are essential aspects of rater-training and rubric-design pilot testing for focusing action-research interventions on groups of rubric raters.

Lovorn & Rezeai (2010) focused instead on writing rubrics, describing two compelling experiments wherein a large number of college students attempt to utilize a scoring rubric to standardize their grading of writing samples. The researchers found the common rubric had no positive effect in this manner, and in fact may have been counterproductive in terms of reliability. Striking a more cautious tone than other authors, Lovorn and Rezaei (2010) posit that raters must be well-trained on an instrument in order to benefit from such a tool, and the rubric itself must be well-designed so as to not encourage holistic and unfocused grading or scoring. The study's intervention drew from these ideas, ensuring that training participants could suggest improvements to the scoring instruments used and enhancing rubrics as SLOA efforts expand in scale to greater numbers of faculty.

Rubric-norming effects as training intervention. Rubric norming belongs to a category of assessment training that has been described by other names in education research: often calibration, consensus moderation, or other terms. Though not identical, these activities all share basic components and theoretical frameworks (Bloxham & Price, 2013; Holmes & Oakleaf, 2013; O'Connell et al., 2016; Saxton, Belanger, & Becker, 2012).

20

Rubric norming components. Though structures may differ, the common elements of rater trainings allow participants to grow assessment-related skill through collaboration and reflective practice (Kogan et al., 2015; Oakleaf, 2009). Reviewing scoring procedures or prescribing certain use of the scoring instrument improves fidelity of rater activity by directly addressing confusing procedural points (Graham et al., 2012; Jonsson, 2014). Many studies have also shown simple review of scoring protocols and rater-bias issues can improve inter-rater reliability (Hansson, Svensson, Strandberg, Troein, & Beckman, 2014; Saxton et al., 2012), and that online trainings can be just as effective as face-to-face trainings (Wolfe, Matthews, & Vickers, 2010). Further, practicing scoring itself improves a rater's confidence (O'Connell et al., 2016), and the experience of defending or discussing rationale behind scoring decisions creates opportunities for critical reflection and subsequently deeper understanding of the assessment activity (Kogan et al., 2015). Bolstering one's confidence in scoring can lead to increased intra-rater reliability, too, as a rater becomes more comfortable with the scoring activity. Overall these practices function as individual professional development while collaboration generates valuable social capital and productivity for faculty members (Reddy, 2011; Turbow & Evener, 2016).

Exemplars and SLOA as socially-constructed. One important difference between different forms of rubric training for increased reliability concerns the use of anchor papers or exemplars. Almost all rater trainings employ sample performances for practice scoring (Jonsson & Svingby, 2007). They are called exemplars or anchor papers, though, when the samples constitute previously-agreed upon standards to which the current pool of raters agrees to calibrate their understanding (Dawson, 2015; Oakleaf, 2009; Sadler,

21

2005). Some research asserts the importance of anchor papers because, without a set standard, raters have no absolute understanding of a learning expectation on the rubric (Rezaei & Lovorn, 2010). It follows then that, with the use of anchor papers or exemplars, the ability of the trained raters to achieve scoring consistency is strengthened.

Because SLOA represents a socially-constructed idea, however, an institution may eschew anchor papers out of concern they preclude raters from exercising expert opinions in the scoring process. The goal of an instrument-specific *norming* session, then, is to increase inter-rater reliability or agreement, building toward consensus on common interpretations rather than absolute agreement. Some scholars suggest that, especially for new raters, the conversations around shared standards is most impactful to better, more reliable scoring (Handley, den Outer, & Price, 2013). This builds on another strand of scholarship asserting that hand-wringing over inter-rater reliability misunderstands the social-construction of SLOA (Bloxham, 2009 Sadler, 2005; Price, O'Donovan, Rust, & Carroll, 2008). Instead, scholars in this vein argue that learning assessment is subjective, and should be, but that collaborative norming among raters still holds value to help constantly anchor and reinvigorate the expert judgments of faculty raters (Bloxham et al., 2016; Price, 2005). The 'middle way' advocated by these researchers informs the creation of this study's intervention and theoretical framework.

Rater bias and workplace-based assessments. A related construct of note is rater bias. In learning assessment situations, faculty raters can exhibit many kinds of leanings that will influence their evaluation of performance (Myford, 2012). Much education research has concluded that, while there are effective methods to helping assessors eschew various biases during evaluation experiences, bias is impossible to

22

completely eradicate (Lovorn & Rezaei, 2011; Rezaei & Lovorn, 2010; Weigle, 1999; Wiseman, 2012). Similar to the discussion around anchor papers and social construction of SLOA, though, the role of bias in rater training is not one-sided.

   Workplace assessments and performance appraisals. Rubric-based assessment in education represents just one form of rater training. More generally, rater trainings focus on orienting a set of evaluators to a scoring situation, a specific scoring instrument, and usually, methods for exercising expert judgment with appropriate amounts of experience and bias (Bernadin & Buckley, 1981; Woehr & Huffcutt, 1994). Because the specific rating task for faculty participants in this study involves students who are near graduation applying job-focused skills in a way that resembles the workplace, relevant research on workplace-based assessments (WBAs) is also instructive to the study. Scholarship on WBAs suggests that educators or mentors in a workplace environment must account for many kinds of sociocultural factors influence performance evaluations in the workplace (Govaerts, Van de Wiel, Schuwirth, Van der Vleuten, & Muijtjens, 2013; Holmboe, Sherbino, Long, Swing, & Frank, 2010). One such factor, professional and personal bias on the part of the evaluator, is entirely normal, cannot be completely rooted out with training, and should not be (Levy & Williams, 2004; Williams, et al 2003). This idea not only dovetails with the social-constructivist paradigm for SLOA (Bloxham, 2009), but it also recalls the sentiment in higher education that adjunct faculty offer valuable insight to students because of their deep professional knowledge. It follows that if rubric-rater training, or norming, completely scrubs faculty members' biases from their judgment, this might actually be counter-productive.

23

Research on WBAs further asserts that rater trainings often fail then because they cannot account sufficiently for raters' *a priori* biases (Govaerts et al, 2013) and should adapt a model that more proactively invites a rater's accumulated influences to harness and minimize their effect (Holmboe et al., 2010; Kogan et al., 2015). This also supports the SLOA-as-socially-constructed viewpoint, advocates of which also contend that consensus-building around learning assessment standards must include faculty's professional experience and expertise (Bloxham et al., 2016; Handley et al., 2013; Price, 2005). Thus, performance assessment is a complex, subjective task for several profound and necessary reasons.

Educators have an obligation, however, to imbue their ratings of student performance with objectivity, to facilitate transparent, impartial, and meaningful learning situations for their students. Training faculty members to use well-designed rubrics is one method for achieving this. How faculty perceive these efforts, and how they perceive SLOA more generally, is the next topic to be explored.

Faculty Perceptions of SLOA

In higher education, greater accountability and consistency in assessment always requires faculty support – they are the ones at the forefront of new initiatives, putting shared principles directly into practice in the classroom. Research on best practices reinforce the importance of faculty in SLOA (Banta & Palomba, 2014), and accordingly, accreditors expect it to occur (Higher Learning Commission, 2016). How does an institution avoid making SLOA an "incomprehensible burden" for its faculty (Schilling & Schilling, 1998, p.63) while still attaining quality and effort in SLOA? Within the study,

24

these questions were explored through faculty perceptions of their responsibility for SLOA.

        Positive faculty perceptions of SLOA. Researchers have shown repeatedly that higher education faculty of all types sincerely desire to help their students excel (Danley-Scott & Scott, 2014; Scott & Danley-Scott, 2015). Faculty welcome opportunities to practice and develop instructional skills in collaboration with their colleagues (Fairbanks et al., 2010; Kezar & Maxey, 2014), and these forms of professional development help instill a mutual level of responsibility and pride among colleagues (Fullan et al., 2015). The effects of this phenomena extend to accountability measures too. Commonly, institutional SLOA initiatives include faculty training programs (Kuh et al., 2015), and researchers have found faculty receptive, unafraid of high expectations from one's institution (Germaine & Spencer, 2016; Rickards, Abromeit, Mentkowski, & Mernitz, 2016). Other studies point to specific components of training to maximize engagement with participants. Faculty expect leaders to demonstrate consistent support for accreditation or other accountability efforts, with sufficient resources and clear communication (Rickards et al., 2016; Schilling & Schilling, 1998). In general, faculty are motivated by transparent, shared goals and learning that is both meaningful and practical (Lyons, 2007; Richardson, 2007; Haviland, 2009). A case study reported by Haviland, Shin, and Turley (2010) demonstrates how this can be done effectively by a college or academic program. The study involved 44 faculty, from an education program in a California college, participating in a set of workshop trainings for a new programmatic assessment initiative. In both survey responses and interviews, the authors found the workshops intervention had a positive effect on attendees. Faculty members

25

reported increased confidence in abilities for assessment work afterward, as well as in their perceived level of support from the college. These participants also indicated that the collaborative element of the sessions benefited them greatly and claimed to have earned a greater understanding of the work to be done, though these effects waned in the months that followed the intervention (Haviland et al., 2010). The example connected to the current study's research design which used faculty perceptions of SLOA to gauge the effectiveness of a training intervention.

Faculty wariness for SLOA training. When perceived purposes of SLOA training do not align with stated goals, higher education faculty exhibit wariness and disengagement with such initiatives (Cain & Hutchings, 2015; Deneen & Boud, 2014). Poor communication or inconsistent support from leadership may obscure otherwise worthwhile efforts, too (Haviland, 2009). Many studies show higher education faculty and staff approach accountability work with skepticism, perceiving extra effort to satisfy accreditors with little local value (Haviland, 2014; Knight, Tait, & Yorke, 2006; Patton, 2015; Rodgers et al., 2013). A fairly comprehensive catalog of faculty ambivalence or wariness can be found in a study at a Hong Kong university by Deneen and Boud (2014). In the study, faculty and staff were asked to make several large-scale changes to their existing assessment practices, and the researchers sought to investigate the forms of resistance these change efforts encountered. Through an analysis of staff dialogues and interviews, the researchers classified multiple forms of resistance and identified forces hindering assessment changes. And although authors reported a respect for SLOA among participants in a general sense, they still encountered much resistance. This pushback fell into three broad categories, from overarching questions about the value of SLO, to

26

practical concerns regarding insufficient time or resources for effective SLOA, to more procedural and normative concerns about how SLOA was implemented around them (Deneen & Boud. 2014). The findings suggest that, even with general agreement among faculty participants, specific perspectives about SLOA and its utility may vary greatly among university colleagues.

Qualitative inquiry for faculty perceptions of SLOA through collaboration.
Qualitative studies have delved deeper into faculty raters who are assessing student work, and how they perceive training efforts for assessment. A form of discourse analysis was used in a qualitative study to explore the ways faculty raters saw their role in a SLOA process and how they interacted with a partner rater (Bullough, 2010). The study used ten teacher-candidate work samples, discussed and scored by four pairs consisting of one tenure-track and one clinical faculty member each. The researcher noted the scoring teams had hewn a compromised meaning of consistency, between absolute agreement and consensus. Instead of explicitly talking about reliability, the subjects applied a large but fairly stable set of rules and strategies to drive their collaborative work and to understand the scoring/rating processes they created together (Bullough, 2010). By exploring how faculty raters build meaning together in an assessment-of-learning workspace, Bullough (2010) offers ideas regarding the design of the current study and its aims: that one should capture data in a manner that remains open to nuanced ideas of consistency put forth by faculty-rater participants.

In another qualitative study of faculty assessing learning outcomes performance with common instruments, Kogan et al. (2015) approached rater perceptions of consistency in a study of faculty leaders from a set of internal medicine residency

27

programs. Through a set of individual interviews and focus groups, participants conveyed that "group consensus on the criteria […] was 'empowering' and helped them feel 'relieved'" (Kogan et al., 2015, p.699). The participants attributed this benevolent consistency to discussions about scoring rationales and exercises to co-construct definitions of terms used in the scoring instrument (Kogan et al., 2015).

This dissertation mimicked the constructivist lens employed by Kogan et al. (2015) for researching rater training situations. The study's intervention brought together faculty with wide-ranging abilities and levels of experience, training them to use a common rubric to assess student work in the classroom. And, as explained above, these socio-historical influences impact faculty raters' perspectives and emerge as participants' perceptions of the rating task are solicited. To further explore how this influence takes shape, collaboration among faculty rubric raters will be described, as a second major sociocultural mediator of SLOA.

Collaboration and Relational Agency

Research affirms faculty inclusion and collaboration as fundamental to sustaining an assessment culture (Bresciani, 2011). Classroom educators interact directly with students, implement SLOA methods, provide key insights for data analysis, and are often tasked with initiating changes or improvements too. In short, they determine whether learning happens at an institution, more so than perhaps any other constituency (Cain & Hutchings, 2015; Ndoye & Parker, 2010). In addition to essential pedagogical interactions, faculty collaboration with colleagues for SLOA was previously shown to catalyze assessment training effectiveness and healthy assessment culture. Faculty participants in other studies, when asked for their perceptions of SLOA efforts and

culture, responded that they desired greater access to colleague interaction and collaborative resources (Rodgers et al, 2013; Schoepp & Bensen, 2016). Through co-construction of learning expectations and a meaningful contest of scoring rationale, norming participants build consistency into the assessment interpretations of a group. How collaboration occurs, and a faculty's role in facilitating it, represents a final mediating factor for consistent, effective SLOA among rubric raters.

Taylor and Robichaud use the term "coorient" to describe collaboration as a sociocultural action: a subject uses another individual to enhance a learning experience or better achieve an intended outcome, and contributes to the other's learning in a similar way (Blacker, 2009, p.32). This conception of collaboration sees activity and learning as a more dynamic, co-constructed process than perhaps other theories like apprenticed learning: interacting professionals learn by challenging and negotiating with one another's interpretation of the activity object (Edwards, 2005). This suggests proactive encouragement for dynamic interaction among participants than in a community of practice or other situated learning context. This aligns well with the epistemological underpinnings of the current study's rubric-norming intervention, which treated the participants as co-learners capable of questioning one another rather than learning unexamined truths passed down by knowledgeable others.

In this study, the base unit of analysis comprised faculty attempting to assess student learning outcomes consistently through professional scoring critiques of student performance. When an individual faculty member collaborates with another in a professional manner, as in the intended research intervention here, the collaborative dynamic is itself a phenomenon. Edwards calls this *relational agency* and defines it as

29

the capacity to envision a collective goal or objective and work with others toward it (2005). She further discusses its importance in activity systems as a driving factor for knowledge creation (Edwards, 2005), and constructs the concept to coincide with the idea that group membership offers reciprocal support for individuals and collaboration for developing collective competence (Paalova, Lipponen, & Hakkarainen, 2004).

Relational agency operates as a sociocultural and socio-historical factor by catalyzing interaction among colleagues, especially when participants bring disparate experiences or expertise to shared work (Edwards, 2011). As a result, in an activity system, individuals realize their outcome is shared, at least partially, by others in the group through interaction. This is partly achieved because, when professionals practice building relational agency, they learn how to react resourcefully amid unexpected occurrences and ask for help in authentic, productive ways (Edwards 2011). "Differences can be seen as a resource in collaborations" (Wright, 2015, p.631), and resourceful practitioners more readily adopt a sense of mutual responsibility and collectively enhance the problem-solving capability of the group (Edwards, 2007). A recent study of Finnish teachers' supports the importance of relational agency, as researchers publicized a link between teachers' activity in collegial or professional circles and their perceived efficacy toward student learning in the classroom (Pietarinen et al., 2016).

This dissertation's intervention purposefully fostered dialogue and interactive assessment activities among participants using the common rubric. Facilitator-encouraged discussion prodded attendees to repeatedly engage their own perspective with others'. This is how shared interpretations grow in a rubric-norming setting, and it demonstrates how relational agency can be a lever to increase both intra- and inter-rater

reliability for faculty rubric-raters: as a catalytic force drawing iterative, critical reflections out of practitioners during collaborative work. Faculty become more practiced, knowledgeable, and familiar with the assessment tasks. These qualities make it a key driver for studies focused on an action research approach (Wright, 2015). This study aimed to capture relational agency as it happens amid workshop dialogue, to identify the role of relational agency in facilitating other intended outcomes of the intervention: increased inter-rater reliability, strengthening of faculty perceptions of SLOA, and support for a developing culture of assessment.

To better understand faculty participants' collaboration in rubric-norming workshops, and how it relates to faculty perceptions of SLOA, the author next shifts attention to a sociocultural learning framework. This theoretical lens will help explain the complex learning tasks occurring within a rubric-norming assessment workshop and locate faculty perceptions and collaboration as specific mediating factors within the framework.

Theoretical Framework: Activity Theory

Sociocultural principles. Seminal learning concepts conceived, tested, and published by Vygotsky and his contemporaries Luria and Leont'ev formed the basis for much of the theory that has fueled education research for more than half a century (Eliam, 2003). Most important is the groundwork laid by these theorists regarding the impact of mediation and social interaction on one's learning. Predominant psychological research at the time largely attributed a person's learning only to brain elements, a subject and the perceived object of learning or action. Vygotsky instead theorized the importance of a subject's interactions with an 'other' as an intermediate influence on activity

31

(Shanahan, 2012). Beyond the human subject and some intended outcome then, a tool or artifact involved can uniquely shape action, creating a complex and dynamic system (Wertsch, 1995). Vygotsky explained that, in constraining or limiting action, the mediating tool "alters the entire flow and structure of mental functions […] by determining the structure of a new instrumental act'" (qtd. in Wertsch, 1995, p.63). Thus, an intervening factor contributes to meaning-making and learning, through its presence and impact on the subject acting toward an intended object (Sannino, Gutierrez, & Daniels, 2009). Figure 1 demonstrates this idea. The subject may be an individual or a group of people with a common aim, though the object is fluid and capable of changing in the minds of individual subjects (Edwards, 2005).
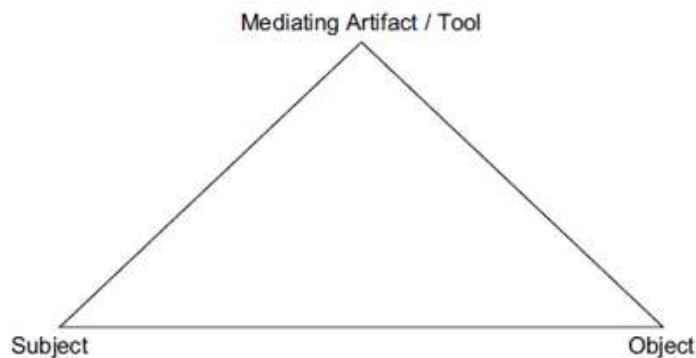
Figure 1. Vygotsky's Basic Mediation Triangle (from Yamagata-Lynch & Haudenschild, 2009)

Socio-cultural and -historical influences on activities. Vygotsky's contemporary, Leont'ev, expanded on the idea of mediated action, formalizing a second iteration of the activity theory concept (Yamagata-Lynch, 2010). Leont'ev proposed that the mediating factor in activity could be another person, or some socio-historical or sociocultural influence on the subject/object exchange (Leont'ev, 1978). Learning or activity then is situated within a constellation of structural, social, and cultural factors

32

such as past experiences, predominant cultural norms, and other constructs (Engeström & Sannino, 2010; Leont'ev, 1978). In more recent decades Engeström has built upon the revised Leont'ev model with a now-ubiquitous triangle diagram, which further organizes and specifies the mediating elements (Engeström, 2001; Sannino et al., 2009), as shown in Figure 2. The Engeström model incorporates categories of mediation: tools, community, division of labor, and rules constitute specific contextual elements of learning. The full diagram is considered an activity system, which serves as the unit of analysis for a study based in activity-theoretical framework. In sum, activity theory purports to explain how knowledge/learning is created and manipulated in complex educational or professional work settings (Yamagata-Lynch, 2010).
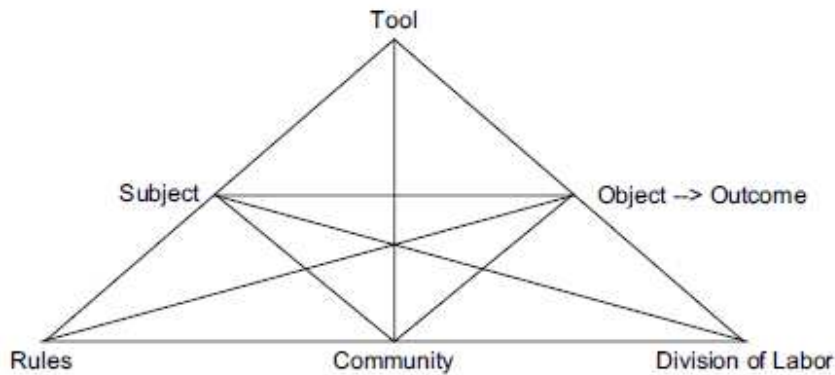
Figure 2. General Activity System Model (from Yamagata-Lynch & Haudenschild, 2009)

Transformation and tensions within activity systems. By capturing the multiple rich phenomena that mediate activity and eschewing a rigid 'subject-to-object' lens (Shanahan, 2012), the activity-theoretical model helps one attain a more authentic meaning of someone else's experience, as subjects themselves understand it (Blackler, 2009). In this way, the theory has been suggested to have a transformative property, helping research subjects isolate, identify, and then overcome barriers in their setting

33

(Edwards, 2009; Sannino et al., 2009). Some activity-theoretical researchers caution that research subjects in restrictive or low-level work settings may struggle to transform practice even when armed with these insights (Engeström & Hannele, 2007; Sannino et al., 2009). At SWU, however, faculty groups participating in cycles of research at SWU were sufficiently empowered to build learning expectations and affect change in their content area and classrooms. In addition, later iterations of action research may affect change on a greater scale by transferring the findings or methods to different settings.

Yamagata-Lynch and Haudenschild (2009) wrote a study embodying this transformative potential while also explaining intra-system tensions as another important aspect of activity systems. When diagrammed and analyzed richly, the components of an activity system often reveal tensions or contradictions inherent in the systems they detail (Yamagata-Lynch & Haudenschild, 2009). This manifests in data patterns that show an issue or repeated conflict; for instance, a system's rules or norms may counteract those of the various community groups participating in a shared activity. In the aforementioned study, researchers spent time understanding how suburban elementary school teachers adopted technology. Though professional development opportunities existed, adoption of the technology often proved unsuccessful because the actions ran counter to pre-existing cultures at the schools studied, or because educators had more pressing needs and could not focus on the technology adoption (Yamagata-Lynch & Haudenschild, 2009). During pilot cycles of research, similar types of conflict were suggested by faculty interactions with colleagues and the alignment of faculty goals with the aims of the institution for specific assessment initiatives. When a study's activity system represents inter-professional collaboration among practitioners, the shared object of activity – the goal or

34

motive – can shift for one or more participants (Edwards, 2005). This was true of the

dissertation study and thus may be the cause of system tensions or issues. Exposing those

tensions through data collection and analysis helped explain their causes and the

subsequent effects it has on the hypothesized goal of the activity system, which in this

case was consistent and effective SLOA by faculty rubric raters in a training

environment.

Activity System as Heuristic for Research Context

The sociocultural principles underpinning activity theory can be summarized as

follows: mediating factors surround complex, goal-oriented learning; these pieces interact

and influence the subject and object; tensions among activity-system components can

explain success or difficulty toward achieving the activity's aim. Modeling a problem of

practice with an activity-system framework can then act as a heuristic, arranging complex

factors so as to classify and explain them. In the study, faculty came to rubric-norming

workshops attempting to build consistency in their SLOA practice as raters, and the

activity-system-as-heuristic-device guided the researcher's understanding of how the

consistency and collaboration is built.

Figure 3 shows the elements of the activity system hypothesized in the current

study. Faculty members approved to teach a specific course entered norming sessions

with the goal of enhancing their consistency and effectiveness with SLOA. They were

part of a community at multiple, ordered levels, within their professional field and inside

the university. Their activity toward consistent SLOA was regulated by the tools they

use, including the norming session protocols and the benchmark assignment and rubric

materials. Further mediation was exercised on the system by 'rules' – by the norms

governing professional and instructional practice, in addition to specific rules imposed by the assessment tasks at hand. Such rules acted as limitations to shape the actions of participants within an anticipated framework. Finally, the actors in the activity system were also influenced by the division of labor or roles within the system. Adjunct faculty members may have particular perceptions of their role amid full-time faculty or staff, while faculty leaders or liaisons impose different tensions or conditions to the system when working with general faculty members. Each of these factors impacted how faculty did consistent assessment work together, and the extent to which the intended outcome – improved culture of assessment and consistent, effective SLOA – was achieved.

Tools/artifacts:
- Benchmark assignment/rubric
- Online classroom and eCampus
- Student learning outcomes
- Norming structure

Subject:
Faculty approved to teach HCS/475 in norming intervention

Object:
Consistent student learning outcomes assessment by faculty

Outcome:
Effective, inclusive culture of assessment at the university as perceived by faculty

Rules/norms:
- Faculty norms
- Norming protocols
- Grading, assessment policies

Community:
University: Program- and college-level faculty member;
Professional field: networks for career and current employment, subject-area or skill expertise

Division of labor:
- Rubric designers and college leaders
- Other faculty who don't share goals
- Cross-collaboration w/staff

Figure 3. Hypothesized Activity System of Faculty-led SLOA within the Study.

Sociocultural factors: interconnectedness and *blurring out*. In an activity-theoretical framework, analysis of sociocultural factors that influence subjects' goal-

36

oriented activity must regard those factors as separate but connected entities. For decades, researchers have been guided by this concept to carefully, thoughtfully develop a study's sociocultural analysis (Foot, 2014; Gutierrez & Rogoff, 2003; Rogoff, 1995). Individual perspectives on a phenomenon of interest must be explored for their sociocultural and sociohistorical roots, in addition to capturing the rich interactions among multiple study participants on these different aspects (Hoffman-Kipp, Artiles, & Lopez-Torres, 2003; Waitoller & Artiles, 2016). When analyzing multiple effects on an activity, the interconnectedness of mediating factors is unavoidable: one's words are likely borrowed and adapted from someone else, and past experiences influence current action in unseen ways (Rogoff, 1995). To avoid compromising the focus of the research, sociocultural studies commonly use a concept called 'blurring out' to more deeply examine related aspects without entirely losing their interconnectedness (Foot, 2014; Rogoff, 1995; Yamagata-Lynch, 2010). The study drew from this approach, exploring two specific sociocultural factors on consistent rubric-norming by faculty while maintaining an appreciative view of these factors' complex connections to the overarching activity system and culture of assessment of the institution.

Faculty perceptions and collaboration as mediators. The dissertation research focused on the mediation of consistent rubric rating through faculty's perceptions of SLOA and collaborative work in rubric-norming sessions. Faculty perceptions were defined previously as comprising personal dispositions toward SLOA, knowledge or understanding of SLOA, and perceived institutional support and leadership for SLOA endeavors. In the activity system model, these perceptions were predicted to operate as norms, influenced by the community and training environment. How faculty understood

37

SLOA would affect their ability to collaborate within a rubric-norming session, and the study's intervention is hypothesized to strengthen these perceptions. In multiple data-collection methods including survey responses, the researcher observed and analyzed changes to these views as faculty participants moved through the study's intervention phases.

The observation and analysis of authentic rubric-norming workshops as SLOA training served as a second medium for capturing another part of the activity system: how faculty participants collaborated with one another in a training environment. This concept was supported by extensive education research literature and guided by the term *relational agency*. Collaboration within the norming training itself was predicted to function as a factor within the activity system, a structure that shapes – and is shaped by – the interaction of the participants. As the intervention prompted participant discussions, those interactions were interpreted using an iterative textual analysis approach. From this process, which is fully explained in the following chapter, resulting themes guided the researcher's understanding of how collaboration is achieved by faculty participants, and how it affected consistent rubric-norming among faculty colleagues.

Summary

Effective SLOA hinges on sufficient inter-rater reliability, represented by the alignment of interpretations by raters using the same scoring instrument. The study's intervention trained faculty to use rubrics more consistently, emphasizing collaboration and critical reflection. In doing so, faculty raters co-constructed the assessment expectations for their area, norming the collective interpretation of the rubric to fit multiple perspectives (Handley et al., 2013; Sadler, 2013). In turn, this strengthened

faculty's perceptions of SLOA and offered insights toward how the institution's culture of assessment can best be characterized at present and improved in the future.

In the activity system model of the current study, faculty workshop participants represented the activity-theoretical subject. The intervention reinforced a set of tasks or operations: applying a common rubric to student work samples, using a standardized scoring process, grounding one's analytic score in textual evidence, and anticipating other perspectives for the same scoring decision. During the workshop and after it, their objective was to construct a normed interpretation of learning outcomes with colleagues. These faculty, along with the organization's leadership around them, also hoped for an eventual, intended outcome of enhancing the culture of assessment within the subject area. A theory-based activity system created a schema for examining whether the goal of consistency in rubric-norming was achieved, and how it illuminated culture of assessment surrounding the study participants through the rubric-norming process. The study also aimed to explain faculty perceptions of SLOA, the collaborative nature of rubric-norming among faculty training participants, and how these two phenomena impacted the achievement of rubric-rating consistency by faculty.

CHAPTER 3

METHODS

In the following sections, I describe the specific context and participants for the study, along with a detailed plan used to conduct the research. Next I explicate my research design and methodology, justifying specific design decisions through their basis in my context and theoretical framework. Finally, data collection and data analysis methods are presented individually with rationale for their inclusion and alignment to the research questions, along with an acknowledgment of study limitations.

Participants and Setting

Academic program and course selection. Academic programs within SWU provide their instructors with a centrally-designed curriculum and resources to facilitate courses. All such facilitation occurs in an online, virtual space, even for the university's physical campus locations that conduct instructional sessions face-to-face in brick-and-mortar buildings. This leads to a hybrid course delivery for the roughly 20% of students who are not exclusively online, while the majority experience all instructional content and course facilitation virtually in the online classroom.

Inside the virtual space, faculty can change content and activities, but courses designated for programmatic assessment data-collection contain some fixed "benchmark assignments" and rubrics. These capstone performances, often at the end of a degree program, prompt students to demonstrate mastery of one or more program-level student learning outcomes through an authentic application of skills or knowledge (Darling-Hammond et al., 2014). As explained in chapter 2, such tasks are most effectively measured using a rubric-type scoring instrument (Banta & Palomba, 2014).

40

Unfortunately, using locked-down instructional measures creates discomfort due to unfamiliarity (O'Connell et al., 2016) and requires training to establish fidelity of implementation (Rezaei & Lovorn, 2010). In this way, all academic areas of the university required training enhancements for faculty teaching benchmark assignment courses. At the time of this research study, the Office of Institutional Effectiveness assessment team had engaged all university schools and colleges to inform about best practices, but university-wide standard practices had not been mandated.

Though SLOA efforts have been initiated in many of the academic programs, the Bachelors of Health Services Administration (BSHA) program, within the College of Health Professions (CHP), was chosen purposefully for the study and merits further description. BSHA contains an expansive set of courses and several program-sequencing options for students. The college's program-level student learning outcomes are grounded in professional competencies, and leaders attempted to standardize these expectations through the common scoring instruments mentioned previously. As a result, BSHA had seen an increased number of assignment rubrics, and a growing number of faculty members required training to use rubrics for both instruction and grading. Also, because the college lacked a pre-existing assessment training structure, the implementation of the intervention did not depart drastically from a previous state and filled a clear need.

Within the BSHA program, study recruitment focused on faculty teaching HCS/475, Leadership and Performance Development. In the default program sequence followed by about 75% of all program graduates, HCS/475 is the final elective course a student takes. It is the penultimate class for the entire program, followed only by a capstone seminar before degree completion. As such, students have accrued almost all of

their required credit hours at the time of this class and are expected to demonstrate mastery of most program-level student learning outcomes through various class assignments. The final week of the course requires students to complete a "summary memo" assignment that synthesizes learning activities from previous weeks of the course and presents a plan to one's leadership for review. This task has practical application and professional significance for the nearly graduated students, condensing a more fulsome explanation of their work into the type of executive summary that may well be required in their workplace. CHP also designated the memo as a 'benchmark assignment', which means it cannot be altered by individual faculty members and must be accompanied by its benchmark-assignment rubric containing nine individual criteria upon which a student's score will be based. The assignment guidelines, rubric, and a student sample of the "summary memo" assignment are included in the appendix.

Faculty recruitment and final participants. Recruitment for the study extended to all active faculty members approved to teach HCS/475; in September 2017 this pool amounted to 26 potential participants. Each approved faculty member was an adjunct-faculty instructor and held some kind of professional position in the health administration field outside the university. A few carried supplementary leadership positions among faculty peers or full-time administrative positions with SWU, heading campus operations or acting as a curricular or student-services expert for the college. These varied ties to the program and university organizational structure highlighted the complex network supporting teaching and learning within CHP and the BSHA program. That complexity hinted at a richness of viewpoints and experiences that faculty would bring to the study.

42

Due to the relatively small size, all 26 from the approved-faculty pool were invited to participate in the study via an email sent to their personal and work email addresses on file with the university. Recruitment efforts resulted in ten consented participants. These faculty members encompassed a mix of professional health care areas, age levels, and years spent teaching at SWU. Each participant brought more than ten years of professional experience to the study, and most had more than 20 years, correlating strongly to their age levels. All ten participants completed both phases of the intervention, though one of the ten did not complete all final data-collection measures. In Table 1, age, experience, geographic location of each participant is displayed.

Table 1

Basic Characteristics of Faculty Participants

Name	Gender	Age Range	Years teaching at SWU	State	Profession
Daniela	F	55+	12	Georgia	Mental health counselor
Erica	F	40-55	2	Missouri	Pharmacist
Gretchen	F	40-55	10	California	Dental Administrator
Nicole	F	55+	8	Pennsylvania	Consultant (Retired nurse)
Vanessa	F	40-55	10	Illinois	Nurse
Sandra	F	55+	13	Michigan	Hospital Administrator
Ella	F	55+	13	Georgia	Consultant (Retired nurse)
Alex	F	40-55	3	Florida	Hospital Administrator
Victoria	F	40-55	2	Illinois	Nurse
Nathan	M	55+	12	Texas	Hospital Administrator

www.manaraa.com

Though the ten represented a spectrum of geographic location, age, and experience within the healthcare field, its representativeness was limited in other senses. Only one male participant signed on to the study although men made up about half of the initial pool. The faculty who responded to recruitment invitations likely differed non-randomly when compared to other faculty groups in terms of responsiveness and other professional traits, although the current study was not been designed to capture these effects. Responding faculty may have possessed increased familiarity with assessment work – or rubric-norming more specifically – or enjoyed established relationships with some of the college or campus leaders described above. The intervention may also have affected the volunteering faculty members differently than non-responding faculty. Given the study design and other practical limitations, though, the partial non-representativeness of the sample was not deemed a substantial threat to the validity of the study.

Action Plan and Intervention

Action Plan. This cycle of action research occurred between August 2017 and January 2018. After preparations made in August, recruitment for study volunteers began in September with an email from the researcher and later from college staff, who normally schedule them for courses. The email included an electronic consent form and two Survey Monkey links: one for formal consent to the study and introductory information, and a second link to the pre-intervention survey measure. Faculty who consented to participate were paid for all data collection and intervention components at a rate of $25/hour. Agreed participants, after consenting and responding to the culture-of-assessment perceptions survey, read a set of four student work samples and scored the four samples using the HCS/475 benchmark assignment rubric. Each of these three pieces

44

was conducted using Survey Monkey in accordance with university guidance. The work samples were scrubbed of personally identifiable information, rendering the scoring exercise 'blind' for the faculty raters. The participants also signed up for a live, virtual norming workshop which constituted phase one of the intervention. The elements of this phase, as well as phase two of the intervention, are explained in the following section.

Table 2

Research Cycle Timetable

Time Frame	Activity
August to mid-September	Finalized study plan with indirect stakeholders and participant recruitment details among those approved to teach course
Late September to early October	Participants responded to pre-survey and scored papers using rubric (data collection point #1)
October	Faculty engaged in live training sessions ongoing with observed rubric discussions (phase 1 of intervention; data collection point #2)
Late October	Participants scored another set of papers using the benchmark assignment rubric and recorded scores (data collection point #3)
November	Faculty participated in asynchronous exchanges on rubric-scoring topics related to SLOA over coure of one month (phase 2 of intervention; data collection point #4)
Late November	Faculty participants took post-intervention survey and scored more papers with rubric (data collection point #5)
December	Share-out of results and preliminary summary with faculty via conference call/presentation expected to occur; shared with indirect stakeholders after (data collection points #6)

Upon completion of all phase-one norming sessions by mid-October, all participants scored a different set of four work samples. Then the asynchronous second phase of the intervention opened, using the online, proprietary SWUConnect networking

45

site. Over the course of a three-week period, faculty were asked to spend at least one hour total time scoring and discussing work samples with the benchmark-assignment rubric, as well as engaging in wider-ranging conversations with colleagues through topics created in the SWUConnect site and a companion Digital Faculty Community site hosted by the college. This second intervention phase was followed by scoring a third set of four student papers and a second administration of the pre-experiment survey. During December, as data analysis spurred initial findings, I conducted a round of member-checking interviews with volunteers from the participants group, to confirm some emerging themes and phenomena observed. This layout is detailed in Table 2, and the relevant recruitment and consent materials are attached in the appendices.

Multi-phase Norming Intervention. The study's intervention was a rubric norming workshop for faculty participants, a literature-based form of rater training described in the preceding chapter. When norming, participants trained to evaluate student work more reliably using common rubrics. This goal was achieved through direct instruction of standardized SLOA processes with a rubric, practical application of the scoring instrument to student artifacts, purposeful interaction with colleagues discussing scores and decision-making, and critical reflection for leveraging norming insights for future scoring tasks. The norming intervention was conducted in two virtual, sequential phases: first a live virtual session conducted via Skype for Business, and then extended-duration access to an on-demand, proprietary networking group site called SWUConnect.

Phase one – live sessions. During the initial data-collection period following recruitment, participants chose to attend one of four live sessions. The maximum number of participants for any single session could accept was seven, and no session was

46

expected to run with fewer than three participants. The researcher reinforced fidelity of the intervention's live workshops by following a structured format and leading each live session himself. Individual sessions still varied in terms of emerging discussion topics or duration of specific workshop activities, but there was sufficient uniformity to ensure all common elements were carried out. Each session lasted approximately 120 minutes in addition to participants' preparation work beforehand. Specifically, participants were paid to do one hour of work prior to norming, which was spent reviewing the rubric, independently scoring four work samples, and jotting down related rationale. The live norming sessions for faculty comprised the following components:

- Orientation to the rubric, benchmark assignment (average duration: 30 minutes)

- Standardized scoring process for analytic rubric (average duration: 15 minutes)

- Discussion of rater bias and other scoring issues (average duration: 15 minutes)

- Review and discussion of scoring rationale among colleagues with pre-scored work (average duration: 60 minutes)

- Practice scoring with the rubric and further discussion of rationale (*additional prepared activity if time allowed*; expected duration: 30 minutes)

These elements represented a mix of direct-instruction topics and collaborative work. In particular, discussion of rubric performance descriptors often led to dialogue around confusing language and substantive focus on what learning meant for each participant. Through the conversation that evolved, faculty had an opportunity to advocate for re-alignment of rubric criteria to assignment task details. They also suggested changes to the rubric performance indicators. This was an explicit example of

47

the co-constructed interpretive activity at the heart of norming, and it allowed faculty to directly improve curriculum and teaching for SWU students. Learning occurred repeatedly, and in more deliberate ways, when the training moved to scoring rationale. As faculty defended a scoring decision among their colleagues, the discovery of others' thought process led to understanding oneself – and one's expert decision-making – as part of the larger faculty base. Inherent in all of this work was a constant, critical reflection on how to more accurately and consistently measure student learning with one's own students. In addition to the "summary memo" benchmark assignment, scoring rubric, and student work sample used in the live intervention, the appendix contains a protocol of prompts used by the researcher when facilitating workshop discussions. The protocol applied specifically to the instrument-focused training. For the purposes of the current research study, some suggested prompts have been included, but routine training exercises rarely allowed for divergence from the framework. The questions were designed to spur practical and reflective conversation around the assessment task in question, which promised a more authentic rendering of faculty perceptions of SLOA and support for SLOA, as well as evidence of relational agency through collaborative work.

   Phase two – asynchronous exchange. In the second, asynchronous phase of the intervention, the format adapted the constructive feedback and dynamic interaction of the live sessions to an on-demand format. The online group site opened at the conclusion of the last live session and remained open for approximately four weeks. Within that time, faculty were invited to join the group via email and paid for at least one hour of specific, active usage of the site. Using instructions posted on the virtual page, participants were asked to read and score two more papers, recording their score and rationale through a

'polling' feature of the site and the associated comments sections. After contributing their own thoughts and work, the faculty were asked to spend additional time reading others' perspectives and engaging one another in asynchronous chats by responding to comments or posts. When another's assessment score and rationale are explained in writing, the next group member could respond and critically reflect on colleagues' thought process. This repeated as the entire cohort engaged individually, developing the collective knowledge base. Overall, the two-phase intervention approach was targeted, methodical, and iterative. The goal was to make each participant a more confident individual scorer while deepening the group's understanding of SLOA and raising collective competence with assessment tasks. The design for the research study is shown below in Figure 4, and is further explained in the Research Design section to follow.

Previous pilot results. In earlier assessment cycles, the assessment team piloted virtual trainings in both live and asynchronous formats, and used different leader/trainer models to deliver the content. No single training mode proved best, however, and constant adjustments were made in response to new wrinkles as they arose. Subsequently, results were uneven but generally pointed to the potential for increasing reliability coefficients when faculty scored student work using common rubrics. Anecdotal evidence also found many university faculty strongly receptive to the training and interactions with colleagues, citing a lack of opportunities for connecting with fellow SWU educators. Finally, previous training pilots captivated the curiosity of the researcher, when observing how faculty used accrued expertise in the subject area to socially construct assessment rating norms. Their professional field experiences, it was

theorized, could provide wide-ranging and authentic insight for how learning skills should be viewed, if applicable to real-world settings.

Research Design and Methodology

To answer its research questions, the study explored the complexities of SLOA and, more specifically, the nuances of why a rubric-norming intervention did or did not achieve its aims. The problem of practice was framed with an activity-theoretical lens. Faculty participants' collective attempt to engage in consistent, effective SLOA was situated heuristically among a constellation of mediating factors, regarded in sum as an activity system. Faculty actions and behaviors while norming to the rubric, and how they collaborated to co-construct learning expectations, was the base unit of analysis for the study.

Research methodology was guided by the activity-systems and sociocultural analysis principles described in Chapter 2, focusing on how the rubric-norming training intervention transformed faculty members' ability to consistently and effectively measure student learning outcomes with benchmark assignment rubrics. These changes were observed through the faculty participants in the study in multiple avenues. Faculty perceptions of SLOA, including content knowledge, personal attitudes toward SLOA, and perceived support/leadership, informed one's beliefs about assessment work and may subsequently have impacted a faculty member's assessment-related actions. The collaborative nature of interactions among faculty during training exercises elicited another layer of sociocultural factors, further affecting rubric-rating consistency among study participants. The interaction of these factors also played a role in faculty participants' ability to apply learned training experiences toward more consistent,

50

Figure 4. Intervention Design for Research Study

Pool of active BSHA faculty approved to teach HCS/475

Data collection point #1
Faculty perceptions survey
Faculty rubric scores
(Pre-intervention)

Intervention Phase 1

Data collection point #2
Faculty rubric/norming discussions – live/virtual
(Within intervention)

Data collection point #3
Faculty rubric scores
(Between intervention phases)

Intervention Phase 2

Data collection point #4
Faculty rubric-norming exchanges – asynchronous
(Within intervention)

Data collection point #5
Faculty perceptions survey
Faculty rubric scores
(Post-intervention)

Data analysis phase: evaluation of college's culture of assessment

Data collection point #6
Member-checking during initial findings presentation
(Post-intervention)

Presentation of study findings to all stakeholders

51

effective SLOA. It was hypothesized that, by examining both sociocultural factors –
perceptions and collaboration – with a mixed-methods approach, the study could show
some correlation between the norming workshops and improved perceptions of SLOA
among faculty, or the inter-rater reliability of their rubric scores.

    Action research methodology. More broadly, exploration of how faculty
collaborated for SLOA also pointed toward bright spots to share with other programs and
colleges at the institution to transform assessment, teaching, and learning. In this lay the
true promise of the study. Mertler (2014) notes that collaborative elements in an action
research study can not only improve local educational practice but may also spur larger-
scale, systemic school improvements. Real change in practice, growing from the
intervention and insights gained from its observation, is an important tie to the study's
action research roots.

    In other ways too, the study was grounded in an action research approach.
Starting in 2014 I began addressing my problem of practice through critical reflection
with colleagues and trial-and-error cycles. I did so without recognizing them as essential
components of action research (Plano-Clark & Cresswell, 2015). The explicit
participation of faculty was essential to my action-oriented approach. SLOA cannot be
understood fully without faculty perspective, nor can interventions or potential solutions
be ventured without connecting to faculty work in classrooms. By enjoining them as
participants in the study, along with the theoretical framework discussed, the faculty and
researcher were ideally situated to navigate and create knowledge from the "dialectical
relationship between theory and practice" (Zuber-Skerritt, 1992, p.11). Following in this

vein, the study's research questions evolved as a joint consideration of factors affecting faculty work and institutional needs for SLOA improvement.

Purposeful faculty involvement also combatted the influence of regulatory/professional standards or top-down mandates from our own university leaders. Taking others' imposed expectations or understanding of SLOA would have amounted to a contrived, post-positivist sense of SLOA's importance for our institution. Instead, by putting faculty at the center, the study welcomed faculty leadership for analyzing data and for suggesting future assessment direction. Faculty thus determined the relative importance of consistency in the collection and analysis of SLOA data and the extent to which their interpretations of a rubric meshed with those determined by other faculty groups at the university. In this way, as Crotty (1998) points out, faculty crafted meaning using already-produced content and then adding their own perspective as well.

Mixed methods case-study design. I employed a case study design for my research, using a single group and a pre-/post-treatment approach. This design specification, stemming from sampling decisions explained earlier as well as practical limitations within my workplace, clearly limited the inferences possible from subsequent data analysis. However, the set-up also offered some practical benefits that made the research more feasible to conduct and its findings potentially more useful to the study participants. The dean of assessment from the college of health professions originally sought assistance with data collection for programmatic accreditation requirements. When the study was introduced to her, she saw value in its use for compliance reporting as well as an innovative method of exploring faculty engagement. Thus, a mutually beneficial arrangement was struck, and her partnership guaranteed a committed

participant base willing to gather data and learn from its findings. Additionally, the college staff assisted with recruitment efforts and may have increased participation among the faculty sub-group chosen.

Integrating to the two methods. Another quirk of the research design informing its methodology was a nested strand of qualitative research within the largely quantitative pre-/post-experiment design. Mixed methods studies have consistently provided education researchers with fertile ground for assessment-specific inquiries in recent years (Guetterman & Mitchell, 2016; Rodgers et al., 2013). Furthermore, sociocultural expert Barbara Rogoff has stated her preference for mixed methods approaches when studying sociocultural phenomena in educational contexts (Glăveanu, 2011). In her estimation, the richness of a research study's context should not be scrubbed away, but the quantitative methods can add to the qualitative for more profound insights (Glăveanu, 2011). Following this idea, this study used a concurrent methodology, converging the quantitative and qualitative data during the analysis phase of the study. Such an approach allowed the researcher to follow up on the trends in survey responses – whether specific items were high or low, comparatively – and slightly adjust the lens by which the qualitative data were analyzed. Similarly, emergent themes from the textual analysis of norming discussions powerfully explained some of the statistical analysis of inter-rater reliability found in faculty assessment scoring. In these ways, an integration of strands in the analysis phase of the study helped converge upon deeper understanding of the problem of practice, especially in answering research question #1 in the following chapter. Put another way, the study design allowed the description developed from the qualitative data to refine and expand on the general findings of the quantitative data

54

(Creswell, 2015; Greene, 2013). As a final benefit, this approach empowered both strands to maximize their discovery power individually. The quantitative data assessed the effectiveness of the intervention while the qualitative data gathered explained participant perspective and negotiation of the experience. In this way, complementarity offered a way to give more fulsome attention to each research question through the various data-collection methods.

Data Collection and Analysis Procedures

Quantitative data. The overview of data collection timeline and methods shown in Table 2 merits further explanation. The use of mixed methods in this action research project resulted in both quantitative and qualitative methods being collected to create complementary strengths (Greene, 2013). Quantitative research methods allowed the study to benefit from standardized data collection where participants were measured in similar ways (Cresswell, 2015). Similar measurements increased the likelihood of fidelity in data collection from all participants. Also the use of multiple quantitative data-collection methods ensured a more comprehensive evaluation of the intervention's effectiveness.

SLOA perceptions survey. Faculty completed a pre- and post-intervention survey to gauge their perceptions of SLOA. The instrument, which is displayed in the appendices below, was adapted from other researchers' recent work around faculty attitudes toward assessment. Each administration of the survey instrument required less than ten minutes for completion by a faculty participant. The survey's psychometric properties were extensively tested and documented in two articles (Guetterman & Mitchell, 2016; Jonson, Thompson, Guetterman & Mitchell, 2016). The survey went

55

through multiple pilot phases and analysis, resulting in 31 items across three general sub-constructs: knowledge about assessment, personal dispositions toward assessment, and perceptions of institutional support for faculty assessment work (Guetterman & Mitchell, 2016). As explained above, these faculty perceptions toward SLOA constitute a fairly comprehensive view of a school's culture of assessment; in this context they helped explain the impact of faculty perceptions of assessment on the norming work undertaken by the participants in the study. With permission from the survey's authors, the instrument was adapted slightly to address the study's specific faculty population.

Analysis of these data was conducted using SPSS, with descriptive statistics offering simple details about individual survey item results among faculty participants. More comprehensively, paired-sample *t*-tests were calculated to analyze changes related to perceptions of SLOA and institutional SLOA efforts. Gains or other effects found when comparing the pre-intervention instance to perceptions at the conclusion of the experiment pointed to intervention benefits. These statistics further illuminated how the assessment-related professional development of study participants was affected by the collaborative work prompted in the norming sessions and how their understanding of SLOA at SWU had been altered during their experiences.

Rubric scoring data. Faculty participants were asked to independently score four student work samples with the rubric before the norming workshops, and after each of the two intervention phases as well. Using previous cycles as a guideline, it was determined that one can thoughtfully read and score four HCS/475 "Summary Memo" assignments in an hour. Thus, the researcher anticipated each data-collection activity with rubric-scoring would require one hour of paid participation from the faculty involved in the study. The

56

amassed scoring data from faculty before and after the training sessions was statistically analyzed for rating consistency using a statistical procedure in SPSS.

The reliability test statistic measures the consistency among scorers using a given instrument. Rather than insisting on absolute agreement for each rubric-rating by all scorers, a norming exercise pushes toward consensus and reliability. Thus, rather than quantifying instances of exact agreement, a correlation coefficient is the most appropriate measurement of consistency (Norman, 2010; Stemler, 2004). Use of correlation coefficients is common in education research to measure reliability of rubric use for assessment purposes (Bresciani et al., 2009; Jonsson & Svingby, 2007; Turbow & Evener, 2015). Because the present study employs more than two raters, though, the specific statistic used was an intra-class correlation coefficient (ICC), mathematically equivalent to a weighted version of Cohen's kappa, (Fleiss & Cohen, 1973; Hallgren, 2013). The statistic produces a coefficient ranging from -1 to 1; the more like-minded the raters are in their judgment of student performance, the higher the coefficient will be, while with 3 or more raters, all negative values indicate the absence of any scoring consistency. This type of quantitative analysis helped determine the effectiveness of norming workshops and paired with survey data to examine how prevailing faculty attitudes toward SLOA were both confirmed and challenged by scoring data analysis.

Qualitative data. Research on raters or scoring tasks, especially those in an educational setting, point to the need for additional qualitative components when studying rater agreement and training effects (Weigle, 1999). This is because, as some authors have noted, inter-rater reliability or agreement statistics present dangerously little information about true impact of trainings, especially through the lens of a participant

57

(Wang, 2010; Weigle, 1999). The researcher employed a set of qualitative data collection methods to address that issue, complementing the survey instrument explained previously. The qualitative component of the study's methodology principally captured the authentic, intervention-based discussions generated by faculty participants, in both formats of the norming workshop conducted. Further, those interpersonal exchanges were thematically analyzed to inform the study's operational understanding of collaboration as relational agency, and how that phenomenon interacted with faculty's perceptions of SLOA. All discussions were recorded at the time of training, then transcribed manually and prepared for textual analysis. During the second, asynchronous phase of the intervention, specific places on the networking site were identified as substantive ground for data collection. These were anticipated to be the comments section of the rubric-criteria polls for each work sample, as well as the discussion board topics created by the faculty participants. In these cases, all text entered was copied into separate electronic documents and prepared for textual analysis as well.

Coding concepts. Once prepared, all gathered qualitative data underwent two iterations of thematic analysis, an interpretive method of rendering the qualitative text into thematic elements using a specified and often structured process (Braun & Clarke, 2006; Plano Clark & Cresswell, 2015). First, the text was coded using a constant-comparative, iterative approach for open, axial, and selective coding (Charmaz, 2005; Cresswell, 2015). The resulting thematic elements were then analyzed with a comparative technique that compared the codes to some of the core sub-constructs by which the study was framed. Referred to as directed content analysis or other labels, this more deductive method compared the selective codes to a pre-arranged set of coding

categories based on the theoretical framework or concepts (Goodnough, 2016; Hsieh & Shannon, 2005; Strand et al., 2015). Many researchers studying learning/assessment from an activity-theoretical perspective employ this method to cohere interpreted themes from analyzed text, but they label analyses as 'grounded' or may not even acknowledge their deductive orientation (Junor Clarke & Fournillier, 2012; Li & Barnard, 2011; Rodgers et al., 2013; Yamagata-Lynch & Haudenschild, 2009; Yamagata-Lynch, 2010). By proactively defining it as such, this study benefits from its stated principles and procedural steps. In two different iterations, the researcher created a coding schema based on the approach, to address two different conceptual pieces tied to the study's research questions.

The first exploration of the text concentrated on characteristics of relational agency, drawing from sociocultural principles in the activity-theoretical framework previously discussed. Gade (2015) utilized an inductive analytical approach to qualitative data gathered in a lengthy partnership with elementary school teachers in Sweden. Her study centered on productive classroom relationships among students and the collaboration between researcher and classroom teachers in action-research scenarios. The resulting articles from the researcher and those building upon her work characterize relational agency as comprising the following elements:

- A student's stated identification of someone else's motivation or perspective.

- A student's identification of a shared objective between two or more students.

- Expression of one's own intrapersonal change as a result of interacting with another's perspective. (Gade, 2015; Wright, 2015)

59

This study hypothesized the same elements may be present in dialogue around assessment work from faculty in the training intervention. Principles of relational agency thus created a deductive mapping to frame the open-coding process. As new thematic elements developed from the textual analysis, the *a priori* traits of relational agency guided the imagining of relationships and patterns among the various codes and code families (Strand et al., 2015).

The second coding process mimicked the analytic approach described above but substituted traits of relational agency for specific elements of activity systems plus the three sub-constructs of assessment culture as defined by the Guetterman & Mitchell (2016) survey instrument. This brand of textual analysis, grounded in activity theory principles and modeled in the work of many current education researchers, can be referred to as "activity systems analysis" (Goodnough, 2016; Yamagata-Lynch, 2010; Yamagata-Lynch & Haudenschild, 2009). During norming discussions and posted asynchronous exchanges among colleagues, faculty members discussed SLOA collaboration, referring to various sociocultural influences on their joint SLOA work. They also expressed their perceptions of assessment-related topics in these channels. A thematic analysis using inductive coding allowed the authentic meaning of the participants to surface, as faculty offered their opinion of assessment as practiced at SWU. Then a second, deductive technique was applied to the coded data as the thematic components of the participants' responses were arranged alongside the activity system heuristic developed earlier. This aligned the emerging code-groups to the theoretical framework, garnering insights as to how faculty perceive various aspects of the hypothesized activity system and how faculty perceptions influenced collaborative SLOA

60

and rubric interpretations among the faculty participants. Using these sociocultural principles and the perceptions-based culture of assessment definition explained in the previous chapter, this analysis provided a tailored lens into some aspects of the institution's culture of assessment as well.

Potential Significance and Threats to Validity

From conception and proposal, this research study offered an avenue of inquiry to the stated problem of practice but also faced specific limitations and potential threats to validity. First, the hope and scale of future studies must be reconciled with the limited reach of the experiment conducted. The researcher chose a single-subject case study experiment as a way to offer the most immediate assistance to the BSHA program and the study's participants, giving insight to their culture of assessment and the specific nature of collaboration among HCS/475 faculty norming groups. Its design and sample size was also limited by feasibility; the researcher could only be of help to the college by executing a simple, short study. By employing a single-group case study design, however, the study could not reveal true causal claims for the norming intervention, and it also fell short of Kezar's (2013) suggestion that researchers move away from case studies and further SLOA research using more comparative-group studies. As another shortcoming of the study design and sample size, it was known that inferences to a greater population of faculty were lost. Greater caution overall was exercised in interpreting and summarizing the results of the experiment. Promising findings and other methodological components of the study may well transfer to other settings; future readers of the work may be inclined to interpret any worthwhile findings as "naturalistic generalizations" (Stake & Trumbull, 1982). In this way, using a lens similar to social

61

constructionism, an audience might be open to contextual authenticity found in the

research.

Table 3

Research questions and data collection/analysis alignment

Research Question	Data Collection Method	Data Analysis Method
To what extent does rubric-rater training improve inter-rater reliability among faculty scoring a student performance assessment in a health administration course?	Faculty participants score sets of work samples using the benchmark-assignment rubric prior to, during, and following the intervention	Statistical analysis (ICC) of rubric scores for measure of consistency among scorers and growth throughout intervention
To what extent does rubric-rater training strengthen faculty perceptions of student learning outcomes assessment?	"Assessment attitudes and knowledge" survey – probing faculty perceptions of SLOA –conducted pre-/post-intervention	Descriptive statistics for individual survey items and scales of items; *t*-tests to compare pre- and post-intervention growth
How do faculty collaborate with one another when assessing student benchmark assignment work in a health administration course?	Observed discussions and discussion posts (during both phases of the intervention): about scoring rubric, scoring process, application to work samples	Directed content analysis approach to coding and thematic analysis for discussions and message board exchanges to blend open coding and relational agency characteristics
How do faculty collaboration and perceptions of assessment mediate their consistent assessment of student learning outcomes?		Directed content analysis approach to coding and thematic analysis for discussions and message board exchanges to blend open coding and activity systems analysis

62

The study also faced an instrumentation threat brought about by a rater-training intervention that did not use anchor papers or exemplars to guide faculty participants' judgment of student work. It was known at the outset, then, being without anchor papers would severely hinder inter-rater reliability among the rater group, and demonstrating positive training effects with ICCs would be even more difficult. Regardless, the decision was made because the set-up more closely aligned to current faculty practice. Additionally, it promoted a social constructionist approach by compelling faculty to further take ownership of processes for norming student learning expectations.

Confirmation bias posed a threat to the research as well. The partnering college (CHP) and the BSHA program, along with the researcher, had a vested interest in finding evidence of strong faculty perceptions of SLOA and culture of assessment within the study's participant pool. Positive results would indicate a solid return on investment for the institution and promising avenues for future research. Because the qualitative component of the study's methods relied on interpretive analysis, a danger existed that the researcher may have unduly tainted the lens with which he viewed the data. To deter such threats, the data analysis phase of the study included processes to enhance trustworthiness. This author planned to conduct the qualitative data analysis using a second coder, each of the pair independently coding samples of the transcribed workshop conversations and comparing results after each iterative stage of coding to check for reliability. Using multiple raters improves trustworthiness of the process, but it also would have required extra people to commit an inordinate amount of time to coding observation data. Because a dual-coder structure could not be guaranteed, other safeguards were pursued. Following the data analysis and interpretation, findings were

63

subjected to member-checking, with multiple readouts and discussions for faculty participants, and opportunities to share with college staff indirectly too. This offered yet another safeguard against unintended bias threatening the validity of the study (Herr & Anderson, 2015). Finally, the complementarity of multiple data-collection and data-analysis methods gave a further validation and protection and check against bias. Mertler (2014) calls this "polyangulation", referring to the validating power of mixed methods and multiple measure for supporting a claim. The comparison of quantitative and qualitative results in the analysis phase of the study strengthened the overall research. It also ensured each of the research questions received comprehensive methodological coverage. To this end, Table 3 summarized the data collection and analysis information from the chapter, and demonstrated the intentional alignment of methods with the study's research questions.

CHAPTER 4

FINDINGS

At SWU, the faculty measure attainment of student learning outcomes using a
common rubric, scoring student performance in the classroom on specific 'benchmark'
assignments. The research study introduced a rubric-norming intervention to faculty
participants, intending to increase inter-rater reliability in their use of the HCS/475
benchmark rubric through interaction with colleagues, practice scoring with the rubric,
and critical reflection on the process. The training also purported to strengthen elements
of the institution's culture of assessment, especially faculty understanding of assessment
and perceived institutional leadership for assessment. These things, the researcher
assumed, make collaborative, effective SLOA possible. The chapter explores how faculty
participants engaged in SLOA amid complex sociocultural factors, with findings framed
by four specific research questions posed first in Chapter 1:

1) To what extent does rubric-rater training improve inter-rater reliability among
   faculty scoring student performance assessments in a health administration
   course?

2) To what extent does rubric-rater training strengthen faculty perceptions of
   student learning outcomes assessment?

3) How do faculty collaborate with one another when assessing student
   performance in a health administration course?

4) How do faculty collaboration and perceptions of assessment mediate their
   consistent assessment of student learning outcomes?

Research Question 1: Rubric Norming and Inter-rater Reliability

The core hypothesis generating the study was that rubric norming could increase the consistency with which a set of faculty colleagues used a specific 'benchmark assignment rubric' to assess student learning outcomes in the classroom. To test these ideas within the sample population of this study, the research analyzed the independently-scored ratings of student work samples by faculty from before and after the study. Statistical tests for significant levels of inter-rater reliability were sought. Overall, low levels of inter-rater reliability existed among faculty scoring with the rubric. Thematic analysis of norming discussions also yielded insights that artifacts of their SLOA work and individual scoring decisions impacted the consistency of faculty rubric-rating and offers ideas as to why the norming intervention did not affect inter-rater reliability as hypothesized.

Finding 1: Low inter-rater reliability correlated to issues within the benchmark assignment and its rubric criteria. The dimensions of the benchmark assignment rubric span nine performance tasks in which students demonstrate learning outcomes attained throughout the course and their entire program. Faculty then rate students' work based on the proficiency in each area: introducing a problem and solution, analyzing the solution and its implementation, analysis of three different aspects of leadership within the problem and solution, conclusion of the memo and discussion of next steps, overall writing quality, and apt use of citations according to APA standards.

Before the study began, all ten faculty participants rated four student work samples using the nine-dimension rubric. All participant ratings are then analyzed for inter-rater reliability, or the extent to which all raters agree on how a given student

66

performance should be scored. Here, the data represent the consistency with which the faculty interpret student attainment of learning outcomes on the HCS/475 benchmark assignment rubric. Table 4 displays a summary of rater consistency for each line of the benchmark assignment rubric, summarized with intra-class correlation coefficients and corresponding significance test statistics.

Table 4

Inter-Rater Reliability, by Rubric Criterion, Pre-Intervention Scoring Round

Rubric Criterion	Intra-class Correlation Coefficient	95% CI Lower Bound	95% CI Upper Bound	F-ratio	*df*1	*df*2
Problem and Solution - Introduction	.386*	.090	.910	7.275	3	27
Analyze Solution	.359*	.075	.902	6.612	3	27
Analyze Solution Implementation	.449*	.131	.927	9.136	3	27
Analyze Leadership Style	.431*	.119	.923	8.568	3	27
Analyze Leader's Role in Conflict	.470*	.146	.932	9.859	3	27
Leader's Role in Effective Workgroups	.346*	.068	.898	6.301	3	27
Key Points and Next Steps - Conclusion	.451*	.133	.928	9.225	3	27
Writing Components	.341*	.065	.896	6.174	3	27
Citations	.254*	.020	.859	4.404	3	27

*$p < .05$; n = 10

The measure of consistency used here, called an intra-class correlation coefficient, is presented in the first column of Table 4, along with the lower and upper bounds of the

confidence interval for the reliability estimate. The final three columns, the F-ratio and degrees of freedom, are the parameters of for the significance testing of the reliability statistic. Based on these data, the rubric criterion "Analyze Leader's Role in Conflict" garnered the highest ICC at .470, and the final rubric criterion "Citations" yielding scores with the lowest overall reliability with an ICC of .254. Each ICC was significant at the $p < .05$ level, but the resulting estimates of reliability could be interpreted as having consistency ranging from "poor" to "fair" based on commonly accepted guidelines to categorize ICC reliability estimates (Hallgren, 2012).

Table 5

Inter-Rater Reliability, by Rubric Criterion, Post-Intervention Scoring Round

Rubric Criterion	Intra-class Correlation Coefficient	95% CI Lower Bound	95% CI Upper Bound	F-ratio	*df*1	*df*2
Problem and Solution - Introduction	.429*	.112	.920	8.243	3	27
Analyze Solution	.209*	.000	.835	3.649	3	27
Analyze Solution Implementation	.473*	.138	.934	9.085	3	24
Analyze Leadership Style	.490*	.160	.937	10.592	3	27
Analyze Leader's Role in Conflict	.380*	.087	.909	7.127	3	27
Leader's Role in Effective Workgroups	.328*	.058	.891	5.89	3	27
Key Points and Next Steps - Conclusion	-.009	-.081	.541	0.91	3	27
Writing Components	.203*	-.013	.835	3.30	3	24
Citations	.610*	.263	.959	16.66	3	27

*$p < .05$; n = 10

68

In this final round of independent scoring by faculty participants, the ten faculty participants rated another set of four work samples using the rubric and a new set of ICC reliability statistics was calculated. As shown in Table 5, ICCs stemming from the post-intervention scoring round generally hovered in the 'poor' and 'fair' categories again. Each of these ICCs was statistically significant at the $p < .05$ level except for "Key Points and Next Steps – Conclusion", which had the lowest reliability statistic at -.009. It was also the only negative ICC calculated in the set. At the other end of the spectrum, the criterion with the greatest inter-rater reliability was "Citations". Its ICC in the post-intervention scoring round was .610, although this same rubric dimension had the lowest reliability rubric criterion based on ratings in the pre-intervention scoring round.
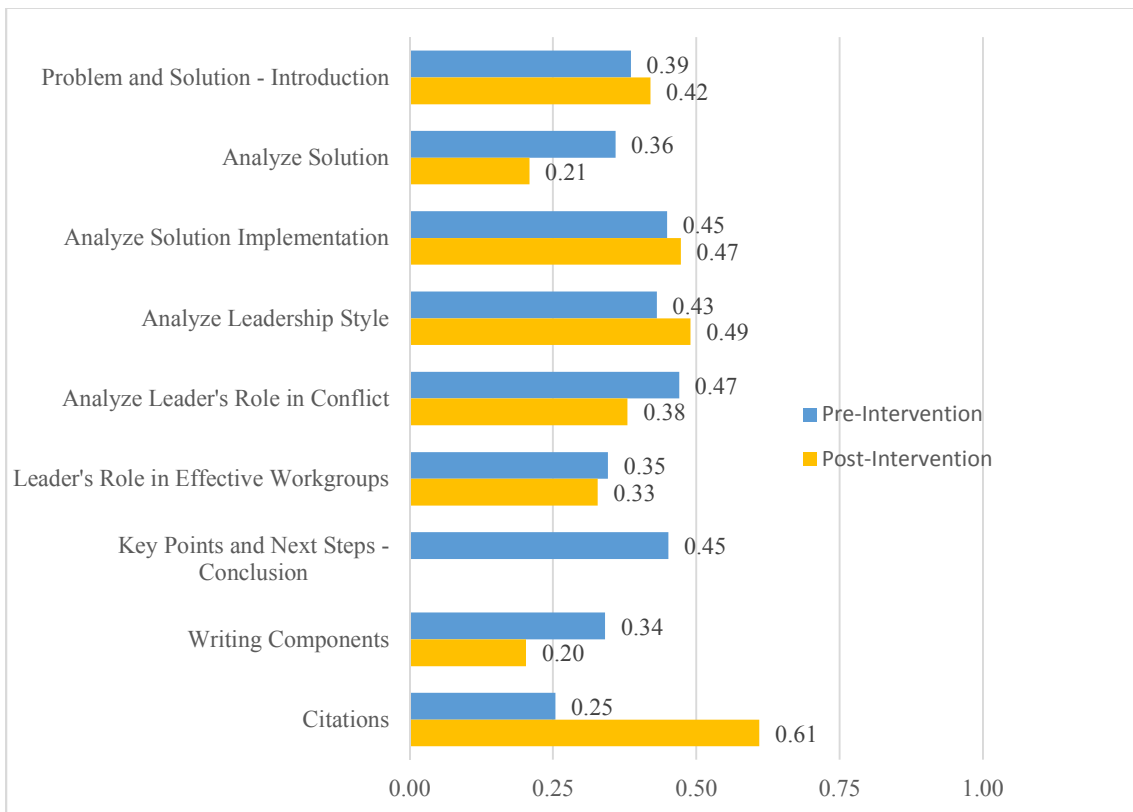

Figure 5. Inter-Rater Reliability, Pre- and Post-Intervention, by Rubric Criterion

Examining each pair of ICCs side-by-side shows a clearer picture of the pattern derived from pre-intervention scoring to post-intervention scoring. In Figure 5, each of the nine dimensions of the benchmark assignment rubric are represented by a pair of lines, representing the difference in consistency among the pre-intervention and post-intervention scoring rounds. Note that Tables 4 and 5 provide the confidence intervals for each coefficient. The study's first research question hypothesized that consistency would increase on rubric dimensions as a result of the norming intervention. However, only three of the nine ICC statistics grew in the post-intervention independent scoring round and none of these differences proved significant.

Rubric language enabling greater scoring consistency. The final rubric dimension, "Citations", assesses students' use of source material following proper APA format. On this dimension, rater consistency increased in the post-intervention scoring round, improving from .25 to .61. The change moved the estimate from a categorization of 'poor' to that of 'good' reliability, according to Hallgren (2012). It was the largest change for any of the rubric criteria, and the final estimate of .61 was the highest single ICC calculated during the study's independent scoring rounds. To determine whether the improvement was statistically significant, Fisher's z' transformation was used to change the coefficients to z-values (Thorndike, 2011). Then a z-test calculated a critical value, but this was found to be statistically insignificant, $z = 1.95$, $p > .05$. Because no other part of the rubric showed substantial increases in inter-rater reliability corresponding to participants' exposure to the norming interventions, additional tests for statistical significance were not needed.

70

The "Citations" line of the rubric merits further mention. Why did ratings of student work became more consistent there than the other eight dimensions? This criterion carried the simplest, most quantifiable language differences, with raters were told to base scoring decisions on number of citations a student used and whether they were used correctly. Research on observer bias has shown that parsing more quantifiable rubric language is easier for consensus-building (Hoyt & Kerns, 1999). This likely affected inter-rater reliability for scoring judgments about student use of citations in the current study, as evidenced by participants' discussions about the criterion. For example, in one live session after hearing her colleagues, Daniela admitted she didn't apply the rubric well previously. She and her colleagues seemed to quickly agree on the proper score:

Daniela: "[A]fter looking at it, she did cite, she did give citations. But the thing about – I was looking at, basically, is that the format wasn't set properly.

Sandra: Yeah, that format wasn't right.

Victoria: Correct – right.

Daniela: I could have given her a 2, but I, I, I was wrong for that one. I was definitely, uh, definitely wrong on that one… [*laughs*]

Victoria: [interjecting] Oh yeah, me too!

One may infer that faculty raters could more easily recall how their norming group had interpreted the "Citations" rubric criterion, applying that information more reliably to subsequent independent scoring rounds. Easier application of shared expectations led to greater inter-rater reliability as compared to other lines with more complex performance descriptors.

71

Rubric language hindering greater scoring consistency. Most other lines of the rubric had structural or language issues that inhibited raters from developing more consistent scoring interpretations during the norming sessions. As shown in Figure 5, the first six lines of the rubric had ICCs ranging from .21 to .49 in the pre- and post-intervention scoring rounds. Examination of the rubric, which can be found in Appendix E, shows these lines share a similar structure. To attain a mark of 'exceeds expectations', the rating scale stipulates a student's performance must have "provided a unique" element, distinguishing it from a performance that only 'meets expectations'. Parsing the word 'unique' – and determining its appropriateness for helping assess student learning – was a topic of much concern to participants.

One participant, Sandra, summarized her group's struggle to develop a shared understanding of what constitutes 'unique': "When I think about it, it's very important to be able to establish some criteria in terms of what it means to 'exceed expectations' because I certainly didn't see that in any of the papers." Victoria, a nurse from Illinois, admitted, "I mean, I could see, easily, how someone would score it a 3 as well. I do. [*laughter*] You know what I mean? This is not unique – because everyone might think, you know, 'What is unique?' You know?" Another nurse, Erica, candidly explained,

> "I will be very honest with you. If they did, if they analyzed it, and I
>
> thought they were a 3, nine times out of ten, I would probably mark them
>
> a 4. [*laughter*] I mean, honestly – if I can't – you know, if you take off
>
> any points, you need to explain it. And so, but it's hard to explain. Ok, so
>
> your perspective wasn't unique: ok. You know?"

Participants in all live sessions mimicked these sentiments, and each norming workshop featured scoring-rationale discussions around "unique" when faculty members parsed the first six lines of the rubric. The issue is representative of the general unevenness in inter-rater reliability among faculty scorers. Even after lengthy discussions in the norming sessions, differences interpreting confusing rubric language persisted, and participants could not achieve greater scoring consistency on those parts of the rubric.

Structure of rubric language impacting scoring inconsistency. In addition to specific verbiage used to distinguish levels of student performance, the way those performance descriptors were structured within the rubric impacted scoring consistency among faculty raters. Sandra, a hospital administrator from Michigan, felt the rubric restricted her ability to give fitting scores to students: "I don't think it lends itself to, kinda, having the necessary flexibility that we need to have sometimes in terms of evaluating a student." Further on in the same session, she said further, "there's just students, I mean, I, I'm measuring them inappropriately, […] but I think that the criteria that are used in how they describe those categories, um, kind of encourages me to do that." In the second quote, Sandra shifts to expressing that the rubric lacked clarity in some of its performance indicators. Thus, following the rubric too closely made certain scores inaccurate and less informative for students who needed constructive feedback. Another hospital administrator, Nathan, had a similar thought in another session, as he attempted to explain his choice of a score-point in rating one student's paper:

> So, that's the main thing is I, um, it doesn't, to me, it doesn't always seem crystal
> clear, um, and […] sometimes it just kinda falls – or maybe it says 3 and they did
> one [referring to a necessary element of the assignment] – it kinda falls between,

73

um, the 'doesn't meet' and 'approaches', or falls between 'meets' and 'approaches'.

Later, he summed up his hang-up with the scoring process thusly, "[W]hen you're trying to move from one to the next, you know, that in-between it gets really fuzzy." A third participant, a nurse named Vanessa, echoed the sentiment of Nathan and Sandra when she said,

> And, um, I can kinda see where there's some gray areas, too, where, um, some of these line-items in the guidelines don't quite meet up with the rubric. The rubric is more, um, it's generalized. It's um, it, there are some specifics in it, but I think if you're going specifically with each line-item with, you know, "summarize the problem and solution", um, specifically, in each category of "does not meet expectations", "approaches expectations" – um, hard to kinda pinpoint and put a, a number on that of how you would assign a value.

In each example, faculty participants admit to a lack of clarity in the performance-descriptor language. Thus, the rubric is one artifact constraining inter-rater reliability within the study's hypothesized activity system. Even as the faculty members voiced their scores and discussed rationale during norming sessions, differences among them for interpreting individual terms in the rubric led to a seemingly intractable amount of inconsistency.

Issues with clarity in benchmark assignment structure. The assignment itself also played a role in rating subjectivity. The design of the "Summary Memo" assignment for week 5 of the HCS/475 course allows students to apply job-focused learning outcomes to a realistic work problem. As a business memo, however, it only asks

74

students for 350-700 words. Whether students could credibly demonstrate all rubric criteria remained a concern in participants' minds. Sandra the hospital administrator asked,

> "[S]o, and, it's only supposed to be 350 [words] – really, how can we in some ways even conclude or, uh, that they have met expectations? When they only have 350 – they only have 350 words – and we want them to be real clear and succinct with it. And it's really, it's, it's, subjective."

In another session, a retired nurse from Georgia named Ella echoed this idea, questioning whether the assignment allowed students sufficient room to exceed expectations: "[I]f they stay within that word count, actually that's just a good 3 or 4 paragraphs, even if it's that much. So, you know, so, I don't know if that's providing that unique perspective or depth of the topic." Later, she continued when discussing a companion resource available to students that instructs how to craft concise business memos. "And in doing so, they might not really express themselves totally, or, you know, provide enough information. Because they're trying to stick to that word limit when they go to that resource." Here the assignment and its parameters for a student's written expression represent a constraining artifact within the activity system. As faculty use its guidelines to coach students through their performance and assess it, they shape how the assignment is interpreted by their class.

Likewise, the assignment description frames the type of work presented to faculty by students, and this impacts how faculty understand the learning outcomes they intend to assess. In this case, faculty participants raised concerns that conciseness was counterproductive for students' ability to fully demonstrate all nine of the rubric criteria.

75

From a sociocultural lens, the tools provided the faculty – the rubric, the assignment structure, and the necessary curricular content – run counter to the intended consistency of rubric scoring within the faculty members' activity system. The qualitative evidence of these issues supports the low levels of statistical inter-rater reliability computed from faculty scoring data.

Finding 2: Effect of individualized rubric-scoring decisions on inter-rater reliability. Norming interactions also revealed many individualized approaches to scoring among faculty participants, highlighting the role unique strategies play in rubric-rating as a form of SLOA. Commonly, participant comments focused on the subjectivity of the rubric-rating task, calling into question whether scoring consistency among faculty was an appropriate goal. Victoria, a nurse from Illinois, found consistent rubric scoring to be a noble but extremely difficult aim: "I agree with you that, you know, that's important that we are trying […] – I don't know how – but, I agree there's definitely a subjective component there, because what I might think there's appropriate might be a little bit different than what someone else does." Nathan the hospital administrator expressed a similar opinion, reflecting on the norming sessions, "It is also interesting to see the variability. This is apparent in classes often as instructors have different requirements. I think it is only natural though to have variability." Gretchen, an administrator at a California-based dental practice, was asked to describe the level of consistency among her faculty colleagues using the benchmark rubric. In response, she put it this way:

> I believe there is unfortunately very little constancy [sic] because the way the
> rubric is designed and also the interpretation of the rubric can be skewed and the

same paper will receive a wide variation of scores depending on how

the instructor utilizes the rubric."

Though use of common rubrics was intended to remove subjectivity from SLOA scoring,

most study participants admitted to employing an individual scoring approach during

assessment. These strategies helped faculty members make SLOA meaningful for their

students, especially in the face of confusing grading policies and problematic curricular

tools.

Table 6

Inter-Rater Reliability, by Work Sample, Pre- and Post-Intervention Scoring Rounds

	Intra-class Correlation Coefficient (ICC)	95% CI – Upper Bound	95% CI – Lower Bound
Work Sample 1	.724*	.51	.91
Work Sample 2	.187*	.04	.53
Work Sample 3	.106*	-.08	.42
Work Sample 4	.180*	.03	.52
Pre-Intervention ICC Average	.299		
Work Sample 11	.386*	.16	.75
Work Sample 12	.244*	.07	.63
Work Sample 13	.150*	.02	.49
Work Sample 14	.370*	.16	.71
Post-Intervention ICC Average	.288		

* $p < .05$.

Statistical evidence of holistic rubric scoring. Most norming sessions treated the

rubric criteria as being linked to one another with overlapping elements. Several rubric

lines shared language or terminology, and participants explained scoring decisions by

referencing earlier judgments and similar thought processes on other criteria. For this reason, the researcher sought to corroborate inter-rater reliability statistics displayed earlier by also treating the estimate as a mean set of nine judgments over four papers, grouped by pre-intervention and post-intervention scoring round. The statistical evidence, found in Table 6, mirrors the unevenness found by the earlier criterion-specific ICCs, with roughly similar levels of consistency found both before and after the intervention.

For each work sample, Table 6 shows the reliability coefficient and a confidence interval for that estimate, which represent the consistency among all the ratings for faculty participants, averaged across all nine rubric dimensions of a particular scoring sample or set. As before, each individual reliability estimate showed raters coming to 'poor' or 'fair' levels of consistency among them. However, a comparison of these numbers also revealed the volatility in the reliability estimates. The very first work sample scored by faculty, when they had been least influenced by participation in the study, produced the most consistent set of ratings across the entire rubric at 0.724. None of the other 13 papers rated during the study evoked such a reliable set of scores, and this score substantially altered the mean coefficient for the first batch of papers scored by participants. The first work sample did not differ substantially from other student performances in writing quality, essay structure, or any other trait. Because of this, the high inter-rater reliability that scores on this sample garnered were treated simply as randomness in the dataset or the result of an unknown, uncontrolled variable.

The first batch of papers overall yielded an average ICC of .299. The average of estimates across work samples stayed roughly the same, however, dropping to .288 on the final batch scored after norming was complete. Some scholars note that keeping inter-

rater reliability at one level over time is itself a difficult task for rater groups (Weigle, 1998), but the impact of the intervention was hypothesized to increase these statistical measures. As such, the norming sessions apparently failed in their aim to develop shared learning expectations and scoring interpretations of faculty raters who participated. Qualitative data suggested that participants may have become more confident in their own scoring judgments over the course of the training, due in part to their reliance on individual approaches to rubric-scoring.

Participant explanations for holistic scoring approaches. In norming sessions, participants often admitted to using non-standard strategies to use the rubric meaningfully with their students. Individual faculty raters explained many different approaches to scoring student work in the classroom during norming discussions. One oft-cited strategy involved a holistic interpretation of rubric elements, due to their similarity in wording and flow, and the perception on the part of faculty that at least some rubric dimensions overlapped in content. As one participant, Erica a pharmacist from Missouri, put it, "In general, I think we kind of grade, period, on writing as one big clump anyways, you know? Um, so, sometimes something else might win or lose over the other…." Indeed, some of her colleagues agreed when discussing their own scoring methods. The retired nurse/consultant, Ella, said, "I kinda just look at the student's overall work, you know, as we're going along, um, on this," while in another session, another faculty member named Alex explained, "that one, I would probably stick with a 3 because they kinda have the whole general nature of it." Combining the rubric elements together represented yet another way that the subjects of this research study attempted to make sense of a difficult, complex task.

79

Even within single lines of the rubric, many faculty touted idiosyncratic scoring decisions. Some raters, like Victoria, the nurse from Illinois, attempted to credit students who gave good effort or were otherwise in need of a boost, by rounding 'up' when scoring: "But they did a great job and really tried, and so I give them that, sort of, leeway when interpreting what they were meaning…." Vanessa, another nurse from Illinois, expressed the same line of thinking as she parsed a troublesome passage in one work sample and tried to award the student for elements not fully realized: "I mean, I get that sentence, but at least they had a bit, I think some good points as far leadership, you know, goes? Um, so I'd probably give them a 3, just to give them the benefit of the doubt and not drag that one down."

Later on, when pressed to explain another score, Vanessa again reverted to this argument, saying, "It's like, when you're right in between, I always want to go higher, just for the student's sake, you know? Just to kinda give 'em the benefit of the doubt? So I'm definitely between a 2 and a 3." This coding theme was prevalent in all norming sessions, especially when scorers faced a decision in between two score points. Of one tricky scoring choice, Gretchen explained, "Ok, I thought it was not quite as low as a 2, but I kinda, I rounded up instead of rounding down. But, ideally, I would like the opportunity for a 2.5, but since there isn't, then, um, we would just have to round down to a 2." Taken together, the numerous individualized approaches undermined the sense that all program faculty share one grading policy or scoring process. Within the hypothesized activity-theoretical lens of the study, the varied strategies interacted with the difficult-to-grasp artifacts to give the impression that faculty cannot obtain higher inter-rater reliability, or do not value scoring consistency. Instead, the study offered a more

80

nuanced understanding of the scenario in which faculty members attempted to norm effectively but were hindered by key sociocultural elements.

Finding 3: Participants' individual scoring rationale stems from a perceived lack of clarity or leadership for adjudicating scoring issues. For SLOA to positively impact teaching and learning, especially using tools like common rubrics or grading processes, communication and leadership is needed to standardize assessment practices and ensure quality. Faculty participants in this study felt this guidance was lacking at SWU, and as a result, consistent rubric-norming among faculty scorers was less likely. One example where centralized leadership lacked was rubric design. As explained previously, several lines of the rubric asked raters to identify "depth" or something "unique" about a student performance to elevate a score from 'meets' to 'exceeds'. Varying the structure of the rubric's criteria, or using more precise language within the descriptors, may have created wider sharing of rubric-score interpretations among participants. Other assessment issues begged for leadership as well.

Scoring vs. grading. At SWU, benchmark assignments rubrics fulfill dual purposes, calculating a classroom grade for the assignment while also measuring student attainment of key skills. Thus, official curriculum guidelines encourage faculty to score student work analytically with the rubric, and only credit students for score levels that they had fully met. Then the online classroom tool automatically computes a total score for the course grade based on rubric inputs. Faculty can override the auto-calculated score to award students more points toward the course outcomes. Unfortunately, this entire scoring process has been poorly communicated to faculty and, as a result, most study participants reported their reluctance to adhere strictly to the prescribed procedure.

81

For some, decisions to not override rubric scores for a student's course grade stemmed from a general discomfort in changing anything. During one session, the nurse Victoria flatly stated, "I was just gonna say, I, uh, I guess I see, I see… I don't ever change it, personally. [*laughs*] I just, I give them what I give them and, and I don't ever change it, the score." However, the laugh seemed nervous, and the multiple stammering revealed a hesitance as to whether her approach was right. Taking off on this same point in another session, another nurse who recently retired, Nicole, was asked if overriding a score was appropriate. She insisted that the student had to earn points solely by demonstrating the learning outcomes, not receiving some form of extra credit: "I don't like the sounds of that. No, I would want them to meet it. The only way I would adjust the grade is if it's late. Otherwise I take the grade it calculates." Her colleague, Gretchen quickly agreed, "Yeah, I don't feel comfortable doing it." She then explained that she also did not change grades manually out of fear that students would not understand and respond negatively:

> And then not only that; um, I think the students would be a little bit upset, um, about that because they're not, they're not gonna fully understand, like, "Ok, how is this the score, but this is the grade?" And I think you would run into, like, maybe people disputing a grade or, you know, issues with that. "[….] So I don't really like that idea, either."

All norming sessions evoked similar comments. Erica, a pharmacist in Missouri, echoed this as she remarked about communicating grade changes to students, "It's just so hard to explain". At another time, Sandra said to her colleagues, "if we have a student who really puts their heart and soul into the assignment and worked very hard, and you say, to

82

'em, you know, 'This is how I assessed you, but that's really all that I expected' – I think it kind of takes the wind out of their sails." These comments carried a connotation of being 'unauthorized' – as in, the faculty did not feel empowered to exercise the change in score. And with so little guidance from their leadership, the risk of causing a hard-to-explain stir among students outweighed the desire to change the score.

When to override an auto-score. At the same time, participants acknowledged that current auto-scoring for grades did not always suffice. As one faculty anonymously posted to the asynchronous site, the norming discussions and training "explained a lot about why when I grade an assignment the score sometimes ends up significantly different." Daniela, the mental health counselor from Georgia, admitted to changing scores when student effort merited a higher score: "Sometimes I would override that, and would give them the, uh, 'exceeds expectations' because they, they have DONE their best, and they have exceeded and they have went above and beyond." In fact, the major complaint in this area of discussions arose as faculty talked of student who completed all aspects of the assignment but only received 75% of the points for having 'met' the various rubric criteria. In one session, Ella, the retired nurse from Georgia, broached the topic by saying, "One thing about the rubric that I have a problem with is that, if they meet expectations on everything, they still don't get full points for that assignment. So I, I have a problem with that. They have to exceed expectations to get, you know, the max points." In another instance, Alex similarly remarked, "it's kind of difficult to ascertain as to whether somebody is going to 'meet expectations', and in my opinion, they did very well and had an 87.5, but yet the rubric is saying that 'meeting expectations' is a 75%. See what I'm saying?" Even colleagues who reported discomfort at changing grades

83

agreed that students not being credited sufficiently for meeting expectations was wrong and should be fixed. In the study's sociocultural framework, these mismatched expectations for grading and assessment created conflict for faculty. Without leadership to determine an official way forward, faculty were left to determine the best course of action individually.

"Nothing is black and white". With several debates over scoring decisions, along with an unclear assessment tool and the natural variety of faculty members' scoring interpretations, many participants expected university or college leaders to exert a stronger hand in guiding SLOA work or adjudicating scoring controversies. Without this direction, faculty defaulted to previous methods of assessment and held little value in rating work consistently with colleagues. Responding to the norming discussions and the apparent lack of consistency among their scoring interpretations in a follow-up email to the researcher, Daniela advised, "Clarifying your expectations for the assignment is an important first step toward creating an effective grading system, one which accurately reflects differences in student performance, lays out clear criteria so that students can gauge their own progress and, most importantly, is efficient, consistent and fair." Victoria made a similar note in an email response by stating, "It is easier to have consistency on the black/white points such as APA formatting, grammar, etc. It is much harder to have consistency with the interpretation of descriptive outcomes." These comments summarized well the breadth of factors impacting faculty raters' consistency in using the rubric for SLOA. Finally, Erica the pharmacist crystallized these sentiments when she stated the following as one norming session wrapped up:

I can't really blame anything for the faculty, unless things that are, like, black-and-white. And I guess, in my experiences with this university, nothing is black-and-white. Like, even as specific as some of the areas are of this rubric, we can give them all 4's all up-and-down and then give them a 5 on the assignment, if we want, you know what I mean? So, it's still, like, there's still room for your interpretation or your adjustment, I guess….

Participant comments like these helped the researcher identify specific causes for the lack of statistical consistency in faculty rubric scoring. The resources used in SLOA scoring – the rubric, the analytic scoring process, classroom grading policies – were unfamiliar or unclear in many places. Worse, the faculty perceive little leadership in clarifying the resulting issues, and in deciding on the 'right' way to complete the assessment tasks. Thus, when left to score student work independently, faculty resorted to 'home-grown' approaches most comfortable to them.

Research Question 2: Faculty Perceptions of SLOA

The second research question challenged the hypothesis that rubric-norming sessions for faculty raters would strengthen their perceptions of SLOA. These perceptions were captured in the dynamic person-to-person interactions of the norming intervention phases as well as a survey instrument delivered to participants before and after the study took place. The survey was adapted from other authors who, in multiple peer-reviewed articles, previously established its reliability and validity (Guetterman & Mitchell, 2016; Jonson et al., 2016). As described in chapter 2, the survey authors define culture of assessment as chiefly comprising three domains of faculty perceptions for SLOA: knowledge of assessment and its uses, personal disposition towards assessment, and

85

perceptions of institutional leadership for assessment and use of assessment data (Guetterman & Mitchell, 2016). Gauging these perceptions and how they shifted as a result of the intervention offers a lens into the nature of our institution's culture of assessment. The data show that faculty perceptions indeed strengthened in specific areas, but survey responses and in-session comments also point to other aspects of the institution's culture of assess where opportunities for growth remain.

  Finding 1: Participation in the study correlated to some increases in faculty perceptions of assessment. Results of the mixed-methods study indicate faculty participants gained greater understanding of the ways in which SLOA impacts SWU, and how faculty members can apply SLOA findings to their own classroom. The first example of this is found upon examining changes in survey scale means among respondents, the results of which are displayed in Table 7. The table displays the t-test statistics for each pair of pre- and post-survey scale means, one pair for each of the survey's three major sub-constructs for faculty perceptions of culture of assessment. Respondents' scale score on the first sub-construct of the survey, knowledge of assessment and its use, grew on the post-study instance of the survey. On a five-point scale where higher scores indicated more positive sentiment or agreement, the average score in this area before the study began was 3.56 (SD = .78) and improved to 4.29 (SD = .69) after the intervention. A paired samples t-test found the overall change of 0.73 to be statistically significant, $t = 2.65$, p = .03. This finding suggests that exposure to colleagues and assessment content during the norming sessions, or simply the awareness of assessment initiatives within the university, correlated to higher mean responses on the corresponding survey construct.

The third construct of the survey focused on perceived institutional encouragement of assessment work. It too brought raw-score gains from pre-study to post-study instances, as shown in Table 7, resulting in a post-survey mean scale score of 3.56 (SD = 1.08). This is the lowest mean score of the three constructs and carried the largest standard deviation, hinting that further improvement can be sought to improve the institution's culture of assessment and normalize its impact on faculty.

Table 7

Survey Response Differences, by Construct, Pre- to Post-Intervention Survey

Construct		Pre-Survey	Post-Survey	m_2-m_1	p-value
Knowledge of Assessment and Use	*Mean*	3.56	4.29		
(Items 1-7)	*SD*	.78	.69	.73	.03*
Personal Disposition toward	*Mean*	4.14	4.01		
Assessment (Items 8-22)	*SD*	.46	.52	-.13	.51
Institutional Encouragement of	*Mean*	3.19	3.56		
Faculty for Assessment (Items 23-31)	*SD*	.97	1.08	.37	.21

Note. n= 10 pre-test, 9-post-test; Five-point scale; higher number equals more positive/greater agreement; *p <.05.

Of the seven items scaled within the assessment-knowledge-and-use construct, three individual survey items saw substantial increases to mean score when comparing pre-survey to post-survey responses, though only one was statistically. These results are displayed in Table 8. Only one item mean difference, regarding faculty knowledge of methods for outcomes assessment, was statistically significant. Growth in this item could result from faculty learning more about how assessment is carried out at universities during the course of the research study's intervention. The other two survey items pertain to collection and use of assessment data for strengthening courses and curriculum. Gains in these items indicate a positive learning experience for faculty by participating in

the study, each potentially improving her depth and breadth of assessment knowledge during norming interactions.

Table 8

Survey Response Differences, by Item, from Pre- to Post-Intervention Survey

Survey Item (Construct)		Pre-Survey	Post-Survey	m_2-m_1	*p*-value
How knowledgeable are you about the methods of outcomes assessment? (Knowledge of Assessment/Use)	*Mean*	3.60	4.4		
	SD	.84	.84	.80	.01*
To what extent have you incorporated the following into your work? *"Systematically collect information about the effectiveness of their teaching beyond end-of-term course evaluations"* (Knowledge of Assessment/Use)	*Mean*	2.78	3.75		
	SD	1.50	1.40	.97	.19
To what extent have you incorporated the following into your work? *"Use assessment findings to inform changes made to their courses"* (Knowledge of Assessment/Use)	*Mean*	3.06	4.03		
	SD	1.67	1.37	.97	.07
To what extent are faculty members at the University encouraged to collaborate with colleagues on improving teaching and learning? (Institutional Encouragement of Faculty for Assessment)	*Mean*	3.61	4.44		
	SD	1.16	.91	.83	.02*

Note. n= 10 pre-test, 9-post-test; Five-point scale; higher number equals more positive/greater agreement; *p <.05.

Within personal dispositions toward assessment, the construct's 11 survey items only garnered one significant mean difference among participants' responses. This item, regarding faculty collaboration with colleagues, had a significantly higher mean score on the post-survey difference and is displayed in Table 8. Statistical improvement in this specific area points to a correlation with faculty's experience in the research study interacting with colleagues and developing a shared sense of student learning outcomes.

88

This supports the evidence and findings in research question #1, based on the idea that faculty perceived a lack of leadership for SLOA at the university. The lone significant gain in this area related to collaboration among faculty for assessment goals, which was very much a focus of the research study. This survey item does not include leadership or other institutional support, however.

Participant explanations of assessment perceptions. Faculty comments during norming sessions support these assertions, as analysis revealed several straight-forward comments wherein participants explained feeling more confident in various aspects of SLOA. Reflecting on progress in an open-box portion of the post-survey, one participant wrote, "I have learned a lot about myself as well as assessment. I find my assessment skills have improved." This comment was mimicked multiple times over by others. During norming discussions, too, participants recognized some improvement came as a result of eschewing individual biases in scoring interpretations. For instance Sandra said, "[N]ow, I've now come to terms with the fact that probably, when I'm grading these types of assignments – and I'm really glad you're doing this – because we do need training." Another participant, Daniela, described a similar sentiment, reveling that she and her colleagues were "not grading a student on personal level and on a level of the student strength and weakness." As will be shown in greater detail through data answering research question #3, working with colleagues and the norming session content afforded faculty participants the opportunity to trade experiences about teaching and assessing, and to transform their own practice through these interactions.

Finding 2: Faculty believe assessment is essential to effective student learning, and this recognition persisted throughout the study. The third survey

www.manaraa.com

construct, personal dispositions toward assessment, probed survey respondents' attitudes toward SLOA and its importance on their classroom instruction. Before the study began, participants' mean score for the construct, was the highest of the three areas and was accompanied by the smallest standard deviation, m = 4.14, SD = .46. In the post-study survey, the construct mean decreased slightly and was surpassed by the assessment-knowledge scale, but its standard deviation remained smallest among the constructs.

Table 9

Top-3 Post-Study Item Mean Scores, "Personal Dispositions toward Assessment"

Survey Item		Pre-Survey	Post-Survey	m_2-m_1	*p*-value
Please describe how you feel about the following statements regarding student assessment at your institution: The effectiveness of teaching is enhanced when faculty regularly engage in student assessment.	*Mean* *SD*	4.57 .73	4.67 .71	-.11	.73
Indicate the extent to which you agree with the following statements: Outcomes assessment would pave the way for better programs for our students.	*Mean* *SD*	4.56 .53	4.44 .73	.11	.68
In your opinion how useful is assessment of student learning to… Improve program or practice?	*Mean* *SD*	4.67 .50	4.44 .53	.22	.45

Note. n= 10 pre-test, 9-post-test; Five-point scale; higher number equals more positive/greater agreement; *p < .05.

This construct consisted of 15 individual survey items, all of which focus on the effectiveness of assessment to (a) support faculty-led instruction of student learning, and (b) to improve academic programs. Examining those within-construct items with the highest mean scores shows specifically the value that faculty participants had for assessment before the study began. The construct's three highest individual item means on the post-survey are displayed in Table 9, along with their pre-test means and a test of

المنارة للاستشارات

www.manaraa.com

significance. No results are statistically significant. Each item shows a slight increase or decrease in mean, but each is nearly at the top of the five-point scale for the survey, with five indicating most positive sentiment or agreement. The results further suggest the study only solidified these specific perceptions of assessment and provided additional aspects or information to enrich the faculty participants' understanding of SLOA.

Exposure to norming and to colleagues' impressions helped strengthen individually held convictions about assessment for faculty. Collaboration often resulted in faculty making statements like this from Alex, the administrator from Florida, at the end of one live session: "No, this was a good experience, and I enjoyed it. I enjoyed hearing from other people that have taught this class too, and, getting your perspectives on everything." Similarly, Nicole noted amid a different live norming, "It's a good discussion; I'm really finding it helpful. And it's making me look at my biases."   In reflection on a session just finished, Victoria from Illinois remarked, "I think it is a great thing to be able to bounce perspectives off of other academic instructors." Indeed, participants' enrichment through hearing colleagues' perspectives was one of the purest and most repeated coding themes found in the qualitative data of the study. This supports earlier statements about the benefits of the norming study for faculty, and echoes recent literature reporting that collaborative rater training is particularly adept at improving scoring confidence and *intra*-rater reliability (Kogan et al, 2015; O'Connell et al., 2016). In the study's activity-theoretical lens, positive experiences developing the normed learning expectations helps solidify the collective goal of consistent, effective SLOA by faculty participants.

Research Question 3: Faculty Collaboration for Assessment

The study's third research question sought to describe the collaboration among faculty participants toward a shared learning expectation during the two-phased rubric-norming intervention. The following characteristics of Edwards's (2005) relational agency construct were used to anchor the analysis of norming discussions, SWConnect online interactions, and other communication with the researcher:

- A student's stated identification of someone else's motivation or perspective.

- A student's identification of a shared objective between two or more students.

- Expression of one's own intrapersonal change as a result of interacting with another's perspective. (Gade, 2015)

In the end, the review process garnered some key takeaways about faculty collaboration toward consistent SLOA.

Finding 1: Frequent, inclusive agreement leads faculty participants to other methods of identifying shared goals and experiences. Small steps toward collaboration occurred in rubric-norming when, after trading views and considering others' perspectives, faculty participants voiced agreement with one another. Most assessment-related assertions by faculty were met with approval by at least one other participant. Frequently, these were simple assents or "*mm-hmms*", and they coalesced toward an informed, shared learning expectation among faculty participants. When a colleague bravely opined about the merits of participating in the study at the start of one session, Victoria, the nurse from Illinois validated the comment, "I agree with you that, you know, that it's important that we are trying." This basic agreement early on helped pave the way for a productive session. In another instance, another nurse named Alex addressed a

92

colleague with, "Um, yeah, Nathan, I'm pretty much on the same page with you on that", and Daniela affirmed a colleague in SWConnect by noting, "Reading your post, I must agree." Over the course of all sessions, these agreements primed participants' ability to accept other perspectives for new angles or aspects to the issue at hand while acting as repetitive identifiers of shared viewpoints in the relational agency framework described above. Within the norming sessions, these instances proved important building blocks to greater collaboration.

Agreement through laughter. Another form of agreement or shared perspective came through laughter in the live norming sessions, often when irksome student behaviors were recalled. In the following exchange, two participants commiserated over bad grammar in student work:

Alex: "And I'm a stickler for those. So, I have to –"

Nathan: "Me too [*laughs*]"

Alex: [*laughs*] "it's really hard to get through some of these papers, you know?!"

In another case, Erica the pharmacist wished aloud – and laughed – that she'd appreciate hearing from even more colleagues, to get their perspective and support as well: "I have not spoken with any more [faculty] […] but, listening to you, like I said, it would have been interesting, because my students act like I grade harder than any professor that's ever walked through the door of life. [*both laugh*]." As stand-alone excerpts, laughter by faculty members during SLOA work could be seen as callous or flippant; and it may have looked even more inappropriate when others joined in on the joke. In context, however, participants often laughed when earnestly discussing messy or ridiculous aspects of their jobs as faculty, as if to signal a shared experience with colleagues. Other participants

93

responded with more laughter, validating their faculty colleagues' perspective and agreeing with the sentiment. This fueled norming discussions, driving participants toward shared objectives and rubric-scoring consistency.

Agreement using shared experiences. A third mechanism for agreement occurred when the faculty referenced common work experience in rubric-norming discussions. As she and her colleagues parsed a student's essay to decide whether to credit for analysis, Victoria remarked to a colleague: "So, the analyze, between you and I, and then Vanessa – you've been a nurse for a while, so we've all been in positions where we are required to analyze down to the 'n-th degree'." In another session, Ella, who began consulting after retiring from nursing, similarly affirmed her colleague with, "And I'm like you, Nicole, so maybe that's the nurse in us." Recalling these experiences signaled a shared value or interpretation, which Edwards (2011) and Gade (2015) identify as another core characteristic of relational agency. In other words, when faculty participants discovered common work experiences with their colleagues, they became more apt at developing a set of shared learning expectations for students, and normed their rubric-scoring interpretations more consistently with others. Thus, using simple agreement, laughter, and shared work experience, the study's faculty participants worked toward productive collaboration.

Finding 2: Differences of opinion lead faculty participants to reflect and change their own practice. Relational agency manifested repeatedly in instances when faculty participants reflected on their own change of mind or perspective as a result of interacting with colleagues in the norming sessions. Often, participants discussed raising or dropping a criterion score based on the explanation they heard from colleagues.

94

Victoria stated to another faculty member, "You bring up a good point. I mean, I could see, easily, how someone would score it a 3 as well." After hearing colleagues explain their lower score of a student in a different session, another nurse from Illinois, Vanessa, said the following:

> I was being very generous to these students! Yeah, no, seeing or hearing, like you said or Victoria said too, going column to column, and, they did talk about – ok, yeah, they're gonna talk about, you know, the issue and blah blah blah – but you're right: they didn't say specifically, "And this what this paper will talk about. This is what we're going to solve", or whatever." So it wasn't really a good connection between the two paragraphs, moving into the "problem and solution". So I definitely would not have given them a 4 after, like, hearing Victoria and Nathan talk.

In the same session, Vanessa and her colleagues exhibited this dynamic multiple times, arriving and new or deeper understanding of rubric criteria or other scoring decisions after recognizing a colleague's viewpoint. After learning why a colleague had given a lower mark than her on another rubric criterion, Vanessa said in response, "You know, Victoria, I didn't even pick that up! …um, so I can see definitely where, you know, that would play into backing it down." Later in the study, on the SWConnect site, the same two colleagues – who also happened to be nurses from Illinois – interacted again when Vanessa said, "Thanks Victoria, I have never throught [*sic*] about using the 1-2-3, A-B-C model. Excellent point - I will try that the next time I teach a class with the benchmark assignment." This realization of another's perspective is a key characteristic of relational agency (Gade, 2015). It also exemplifies the type of critical reflection that can make

95

rubric norming exercises – or other collaborative exercises – so powerful and potentially transformative (Holmes & Oakleaf, 2013). Though norming sessions did not increase inter-rater reliability among the participants, evidence of critical reflection and transformation of practice suggest collaboration generates other positive phenomena of rubric norming for faculty participants working toward consistent SLOA with colleagues.

Recognition of personal bias. In a more introspective turn, participants also talked about how the scoring rationale of colleagues revealed their own biases. One participant, Nicole, remarked during a session to her colleague, "Actually Ella, you're making me rethink some things, so I'm enjoying this. Maybe I am too hard!" Later, in a follow-up email to the researcher, the retired nurse reflected:

> I left the meeting feeling biased and very stubborn [b]ecause of one issue I was
> very hard on the students. This was based on my role in staffing and education
> and the impact of turnover on outcomes. I apologize to the group for not being as
> open minded and fair as I should have been.

In another session Daniela realized she had given credit for performance not really supported by facts. The mental health professional from Georgia explained, "[B]y listening to what they're saying, you know, [the student] did not address, now, the reason why I gave it a 4? But I could back it up to a 'met expectations'. Because, um, enough information is not there." These remarks point to the impact that collaboration and critical reflection have on one's own scoring process. This again links to a key characteristic of relational agency and reiterates a literature-based theme essential to rubric norming. When faculty are provided an opportunity to critically reflect on practice, the collaborative effects of assessment rater training are strengthened (Kogan et al., 2015;

96

Jonsson & Svingby, 2007). Because these elements also make up several key pillars of assessment culture (Guetterman & Mitchell, 2016), the camaraderie of norming exercises may also be considered as creating a healthier overall culture of assessment.

During the asynchronous portions of the norming, there were fewer explicit statements of agreement. Rather, in the scoring and the close mirroring of comments and rationale after scoring student performance in the group webpage, the fruits of relational agency became more apparent. Figure 6 shows a thread of scoring comments building upon one another, displaying the extent to which their rationale borrows from others' thoughts. Even in mimicry, this is further evidence that faculty used collaboration during the research study to bolster their SLOA prowess by interacting with colleagues and being influenced by their contributions. Through all of these sessions and opportunities for scoring practice, the evidence points largely to all participants voicing increased or renewed confidence in their ability to score students and accurately assess their attainment of important, job-related learning outcomes.
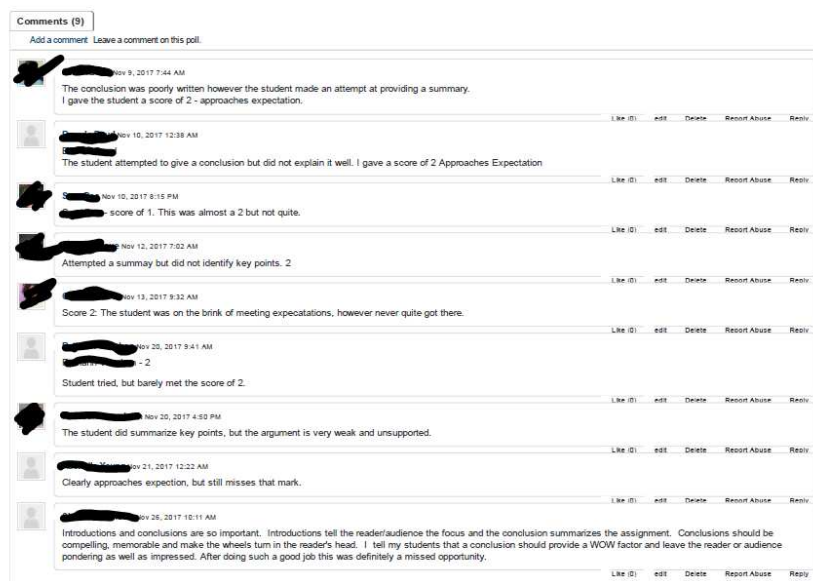
Figure 6. Screen-Capture of Asynchronous Scoring Discussion among Participants

97

Finding 3: Faculty participants identify points of dissent but do not explicitly create debate or argument when differences emerge. Another quality of collaborative ability abounded in the research study when faculty participants confirmed the merit of another's perspective – saying "good point" – as well as the turn of phrase "I can see how/why" when referring to another's point of view. During one session, Sandra, a hospital administrator, asserted the health care industry is founded on documentation and evidence, and that this should be central to the performance assessments the university asks of its students. Alex, a fellow administrator from Florida, followed with, "Yeah; you make really good points, I think, with regard to evidence. And, and, you're right – everything in medicine has to be evidence-based: 'show me the science', right?" At another point, Erica demonstrated this form of validation by responding to a comment: "I understand. No, I get what you're saying, 1000%." Instead of identifying a shared goal, Alex and Erica here endorsed their colleague's perspective. This characteristic of relational agency, whereby participants may not agree but confirm the value of multiple opinions, credits another faculty member for a valuable contribution to the rubric-norming discussion. In these comments, faculty demonstrate respect for colleagues' expertise and suggest that consistent SLOA requires an appreciation for other raters' scoring perspectives.

Deference for others' perspective. Respect for one's colleagues led faculty participants to show extreme deference to others, even when pronounced rubric-scoring differences were uncovered during norming discussions. Participants were loath to press another faculty member on their stance, or to engage in any sort of debate about whose perspective might be more accurate than another; instead of challenging, new ideas were

98

almost purely additive. In one example, Victoria downplayed the difference between herself and another colleague in scoring: "So, yeah, that difference between 3 and 4 is minimal. I don't think there's a huge difference there, so I can see why our biases might just lean us one way or the other." In another instance, Nicole told her colleague, "See now, I'm a little bit different. I do have a lot of students where English is their second language, uh, but when they're at the 400 level […] I guess, I have a higher expectation, the higher course is." While pointing out a significant difference in their interpretation, the retired nurse avoided speaking about her colleague and concentrated on her own rationale. In response, her colleague Gretchen simply responded with "Right", and the discussion moved past the discrepancy. Later in the same session, Gretchen the dental administrator told the group, "So, the, I actually thought the opposite when I saw, at the end…", but again her colleagues did not address the disagreement.

Facilitators of rubric norming, as the researcher in this instance, commonly probe differences among raters and prompt specific participants to explain scoring rationale (Holmes & Oakleaf, 2013). Here, however, these interactions were blunted by faculty participants avoiding any confrontation with their colleagues. This represented a structural limitation within the study's hypothesized activity system, restricting the consistent SLOA by faculty participants. Although collaboration generated exposure to shared goals and an appreciation for others' perspectives, norming the group's collective scoring interpretations seemed empty without fostering debate. How could participants' agreement and relational agency within a training not translate to independent scoring immediately after the session? As an alternative lens by which to explore this question, I was guided to an adjacent field of research in social psychology, on conformity.

99

Collaboration effects as conformity. Early conformity research is highlighted most famously by the Asch studies of the 1950s, which found that individuals were sometimes influenced by group norms to knowingly corroborate inaccurate information. In the years that followed, however, analyses of Asch's findings often lacked the moral and social nuance they deserved, and more recent scholarship commonly sees these seminal works as misinterpreted (Friend, Rafferty, & Bramel, 1990; Hodges & Geyer, 2006). Researchers now offer alternative explanations for group-member behavior, describing conformity within a values-pragmatics paradigm that manages the complexities of social agreement and integrates them into one's established personal truth (Hodges & Geyer, 2006). Across situations or demographics, people develop truth-telling or avoidance strategies for conversation and group work based on a sophisticated calculus of factors (Hodges, 2014; 2017; Mills, 2014). These approaches help individuals conform to the perceived norms of a group, fitting one's own behavior within the bounds of what is likely to be accepted.

In this light, nearly all collaborative mechanisms for agreement described in the previous chapter can also be framed as attempts to conform to perceived group norms. When Nicole referenced shared nursing experience with her fellow retired colleague Ella, she was reminding the group of a shared identity. When Alex offset her admission of overly-generous scoring with laughter – and when her colleagues laughed in response – the participants were signaling acceptance to one another and reinforcing group norms or motivations. And when faculty members specifically state, "I can see that" to one another, they are validating an opinion from the group's perspective. Members use this information to continually triangulate their position within the group, discerning the

100

extent to which they conform to group values and whether dissent is acceptable (Hodges, 2017). Using contextual cues and historical information, the collaborative dialogue amounts to an elaborate dance together. This lens was applied by Trace et al. (2016) to a pair of faculty raters negotiated rubric interpretations together to arrive at consensus assessment scores. In the observed dialogue, researchers found that the two faculty members took care to express their own opinions while validating the other's values, fitting both perspectives into a fuller understanding of the shared learning expectation.

In the context of the current study, conformity research offers an elaboration on the sociocultural factors found in the online discussions of faculty colleagues during rubric-norming sessions. As participants draw from common academic and professional group membership, along with their personal attitude toward SLOA, their mediation of sociocultural factors toward creation of shared learning expectations can be seen as attempts at conformity as well.

Research Question 4: Impact of Collaboration and Faculty Perceptions on SLOA

Faculty members in HCS/475 navigate a complex and competing set of sociocultural factors when teaching and assessing student learning in the classroom. The study's faculty participants evinced these contextual factors through collaborative assessment scoring and discussion in rubric-norming sessions. The activity-theoretical framework for the study includes an activity system heuristic, first described in Chapter 2, which organizes the factors that surround and impact participants' attempt at consistent SLOA. Using the diagram/heuristic and a specific form of directed content analysis sometimes referred to as activity systems analysis (Yamagata-Lynch, 2010), the researcher explored how faculty perceptions for SLOA and collaboration interacted with

101

other competing factors as participants sought rating consistency and normed their interpretations of the benchmark rubric. Analysis found that participants identified strongly as faculty practitioners, wielding dual academic and professional mentoring roles to effectively teach and assess students. They held strongly positive perceptions of assessment and valued SLOA collaboration, but these elements did not translate to higher inter-rater reliability among participants using the benchmark rubric. Instead, they found the most meaningful assessment derived from their own often-biased rubric interpretations rather than a communal understanding of the scoring tool. The following sections explain how faculty perceptions and collaboration operate within the activity system to impact rating consistency, and how participants describe their faculty-practitioner role as influenced by those constructs and the activity system as a whole.

Finding 1: SLOA's importance pushes faculty to academic 'gatekeeper' role for students nearing graduation. From the outset, faculty participants in the study registered some strongly positive perceptions of SLOA. One aspect of these perceptions was the importance of SLOA for teaching and learning; throughout the study, faculty participants showed a sincere desire to facilitate learning for students in the final steps of their journey toward a degree.

Helping students learn with and without the rubric. In class with students, faculty participants credited rubrics with clarifying how skills would be assessed for course outcomes, making both instruction and assessment easier. During one session, Erica admitted, "Um, honestly, if I had the time, and the energy, to make all my rubrics this detailed, um, I would." In another live session, Sandra explained how the rubric helped her prepare students to be assessed "[B]efore assignments, I'd say, 'Let's go over'

102

– I referred to it as the assignment specifications – 'what the instructor is looking for'. I think the rubrics are great." In the same discussion, her colleague Daniela, a mental health counselor, agreed and added,

> So it's like a, like my piece of mind so that, when they say I'm scoring them wrong, and I can show them on their paper where they did not, did not analyze it, or did not talk about what they, uh, need to do, or they didn't have the references or citations, then I can show you what you did wrong.

During an online message-board interaction, Alex told a colleague, "Rubrics give structure to observations. […] Instead of judging the performance, the rubric describes the performance. The classroom rubric is easier for the students to understand because the assessment rubric provides a performance base…." Following a live norming session, I asked one participant, Daniela, what she saw as the overarching value of the rubric we were using. She replied it "can help reconcile the instructor and student perceptions about grades and help students learn to evaluate their own work according to these standards." Leveraging rubrics like this, to review learning expectations and increase assessment transparency, is a desired outcome found throughout the education research literature (Jonsson, 2014; Reddy & Andrade, 2010). These comments show the faculty tried to use the benchmark assignment rubric, and its breakdown of outcomes into performance tasks, as a way to foster transparent, meaningful assessment. In some cases, they were successful in doing so, fulfilling their activity-theoretical aim of doing effective SLOA that positively impacts students.

Faculty efforts in spite of rubrics. In other cases, however, the rubric proved counterproductive to such efforts. Gretchen the dental administrator described her

attempts at helping lower-level writers to improve their skills despite the rubric's rigid scoring structure:

> That's the, the one part of the rubric where I do struggle, because that – I feel like, you know, I'd rather help them. I do want to show them what the errors are […] And I know that, by the end of the course with me doing that, and returning all the papers like that, they look at my changes and they grow and they build and they improve their writing, um, based on what they see and what I fixed in the paper and everything. So even though, it dep – the actual paper they get back looks harsh because half of it may be red, because of spelling and grammar – I don't tend to take off the full 10% for that, because I feel like, I don't want to scare them, I don't want to, you know, make them not be confident in their writing. And I tell them up front that that's what I'm going to do, so they can actually grow and improve on their writing.

This explanation not only highlights the care Gretchen puts into providing feedback to students on their writing, but how she worked to circumvent a rubric-scoring set-up that seemed too punitive and not sufficiently constructive. Ella detailed similar attempts to manipulate the online rubric's auto-score feature in service of accurate feedback to students:

> If I think they deserve a 3, or whatever, and the score came out a little less than that, I don't have a problem changing it and putting a comment in. Um, or even if I lower it, or whatever, putting a comment in, "This would have been more effective had you included this or that." Um, you know, or, "You did a very good job here because of this or that."

104

Finally, Sandra, a hospital administrator, put it more simply in a different session by stating, "Thank goodness that we still were permitted to give a grade not entirely based upon this." The online tool's auto-score represents yet another impediment to effective SLOA for faculty participants in the research study's activity system. Some participants indicated discomfort with changing rubric scores, while other participants like Ella and Gretchen enjoyed the flexibility in assessment scoring, especially so late in students' program sequence. In short, due to the perceived importance of SLOA in the classroom, faculty raters valued giving clear, prompt feedback to students in whatever style they deemed most effective.

Personal and professional biases in SLOA scoring decisions. Because they perceived poor communication regarding standard classroom assessment practices, most faculty members felt comfortable with bias and applied their own mix of expertise and experience during collaborative rubric-norming exercises. Multiple participants acknowledged inordinately valuing a specific element of the performance task or rubric because of past experience or opinion. In one case, Gretchen explained a scoring decision thusly: "I had to step out of, you know, being biased in order to give them a 'met expectations,' because of something else that wasn't related to that line on the rubric." This admission of bias was mimicked by several other participants throughout the study; to best adhere to the rubric as intended, raters tried to consciously check their own scoring decisions had been skewed by biased reading of the work. In the same session Nicole was asked why she had rated one criterion lower than her colleagues. The retired nurse admitted, "I kept looking for something about retention and recruitment, and I wasn't finding that[....] Um, but I just, recruitment and retention was so important in

105

nursing that I, I tend to go overboard on that. That's one of my biases. I have to pull back." In a way, this explicit mention of biased scoring typifies the interplay of sociocultural principles underpinning the study: professional experience interacted with a participant's reading of the rubric and she knew this would affect how her score cohered with that of her colleagues.

During norming sessions, faculty participants validated each other's biased scoring rationale, using it to enrich their own understanding of learning outcomes and potentially transform practice. Despite this apparent open-mindedness in sessions with colleagues, independent scoring among raters did not become more consistent after the norming sessions. After being exposed to colleagues' bias, instead of referencing their group's shared learning expectations, individuals defaulted to an understanding of SLOA that made sense to themselves and their students. Erica, the pharmacist in Missouri, explained, "I just, I don't know what the expectations were before […] so, you know, I feel like I should grade accordingly. And if that forces you not to complete or pass this course, well, I'm sorry." Nicole echoed this sense of raised expectations and feeling apart from one's colleagues when she noted a scoring difference in the session: "See now, I'm a little bit different…. I guess, I have a higher expectation, the higher the course is." Participants regarded all assessment as important, but these comments revealed the separation among colleagues and the stoic responsibility some faculty participants felt during SLOA tasks. From an activity-theoretical lens, this dynamic acted upon the common goal of consistent, effective SLOA. The perceived differences among participants using the benchmark rubric made scoring consistency harder to achieve and less important as a shared goal for the group. Their collaboration and perceptions of

106

SLOA led them to less concern about inter-rater reliability; instead, these adherences to personal conviction allowed them to levy a more authentic and meaningful assessment to their students.

> ***Maintaining standards of learning expectations – the academic gatekeeper.***

Norming discussions, as participants ventured toward a collaborative understanding of the rubric, revealed another root of the individualized SLOA approach espoused by several faculty participants. Because HCS/475 often represents a student's second-to-last course in the program, course facilitators feel an obligation to render summative assessment decisions and hold students accountable for assumed and written standards of performance. One participant, Erica, confided, "[S]ome of the writing that I get at this late in the game from the students, it's concerning." In another session Vanessa, an Illinois nurse, summarized her expectations of students, "I would think at this level, they should be a little, you know, some students should be a little further along than what they at least are portraying in their work." Later in that same session, Alex described why she so adamantly expected students to use a certain writing resource to enhance their essays and other work, saying, "not only is this Week 5, but if this is the last elective class when some of them are actually toward graduating, then yeah, you should be able to, uh, use that [tool]." All three examples demonstrate an expectation that one's learning standard should not be negotiable, regardless of how other faculty members would see the student performance. This reiterated the idea that individual scoring decisions exemplified the conflict in the research study's activity system, as faculty participants uncovered misalignment between others' goals and theirs. Such unevenness made it more difficult for the faculty to achieve consistent, shared SLOA practice.

107

Finding 2: The nearness of graduation pushes faculty participants to act as professional mentors for students entering the healthcare field. Faculty care about students' attainment of specific learning outcomes in the classroom, but interactions from norming discussions suggest the true measure of success is successful application of these skills to the workplace. Their focus on students' career readiness reflects how important faculty perceive SLOA to be, especially in a course so near program completion. Because, as shown earlier, not all faculty members share the same understanding of the benchmark assignment rubric, study participants exited collaborative norming work still worried about personal accountability for student preparation for jobs, and how that preparedness reflected on the university.

Translating the classroom to the workplace. To build proper learning expectations in this penultimate course for students, faculty participants tried to marry high academic standards with professional expertise – what competence looks like within the healthcare industry. While teaching HCS/475, faculty members push all students to continually develop abilities relevant to the workplace; in response, the faculty recalibrate how they assess those job-related skills. One participant, the Michigan hospital administrator named Sandra, described the attitude underlying this approach when she remarked, "Because I tell my students, you should not have to wait until you graduate for individuals to, for your boss to see that your skills have improved." Alex explained how this extended to students new to her professional field as well:

> I also take into consideration the student's background, if they are in health care, if they're not in health care. I have some students that have absolutely zero experience in health care, and they apply what they're learning very well, and

108

they write very well, and they 'get it'. They get all the concepts. And, they do fabulous in the class.

In another session, Nicole, the Philadelphia-based retired nurse, said of her students entering the workplace, "I don't expect them to excel, but I expect them to be prepared." Regardless of student ability level, faculty in the study appeared ready to facilitate job-relevant learning. This represents an important connection between the community ties and professional influences that influence how faculty interact with students, and how they prepare them for the healthcare field during their final coursework.

While focusing on classroom performance and learning outcomes, HCS/475 faculty members also kept students thinking about how those learned skills would apply to a workplace setting after graduation. During one norming session, participants discussed how they prepared students for the culminating benchmark assignment. Sandra stated the matter simply to her colleagues, "…one of the things that I do with all my classes is that, you know, after making sure that I meet all of the course objectives and then to say, 'How do you apply these in the workplace?'" In the same way, another participant, Nathan, the hospital administrator in Texas, explained how he provides students assessment feedback based on professional expectations:

[W]hat I try to do with the students is, you know, I say, "Yeah, you know, technically you, you did get an A on this, your APA was good, your structure was good, you had all of the elements in there. But let me tell you: if, if I was your CEO or I was your boss, and you came to me and asked for $50,000 to implement this solution, um, I wouldn't give it to you. And here's why." And sometimes that's punctuation and grammar. And I, you know, I try to, um, impress on them

109

how important that is. Because, if you're not detailed enough on your grammar and your punctuation, and explaining why you feel or you think that such-and-such is a good solution, then why would I want to give you $50,000? And they usually get that explanation.

As they converted work experience into a meaningful SLOA approach in the classroom, faculty members in this study demonstrated how their professional-community ties act as a substantial influence on the activity system – and the overall goal to achieve consistent, effective SLOA with their colleagues.

Professional competencies. Showing how to use specific competencies on the job represented another specific way faculty mentored near-graduates. In the study, participants explicitly mentioned helping students identify applications for critical thinking skills in the workplace, and the value of that translation. In a message board discussion Daniela, a mental health professional explained, "I tell students that developing this skill will promote them in the workplace and cause them to be noticed. Then their ears really perk up." During a live session, Alex reiterated this point, saying of critical-thinking skill development, "That's important. If they are to advance and go into grad school, they will have a lot more on their plate with their research. It starts with the building blocks here." It stands to reason students understood how to apply relevant content knowledge or specific technical skills, but the professional expertise of experienced faculty members helped with a more abstract skill like critical thinking. The finding reinforces the importance of faculty developing professional expertise, and how that can impact effective SLOA in the classroom.

110

In participants' view, the benchmark assignment and its rubric also influenced how students translated classroom skills to the workplace. Several benchmark rubric criteria distinguish performance that 'exceeds expectations' if the student performs the task "with a unique perspective", while another criterion called out a "unique or creative summary" as denoting the highest-scored performance. Sandra, a hospital administrator, notably questioned this language:

> [I]f someone came to me and said, "This is, I described it in a unique and creative way", it's like, this ain't art class! [… A]nd so, we have to remember that we are all products of the industry that we work in. And so I wouldn't necessarily be looking for unique and creative.

Sandra's colleagues in the session agreed with her. In another norming group, Victoria commented on the difficulty of assessing students' skill attainment with the HCS/475 rubric:

> I think it is, only because I think it's pretty easy [...] to write a paper, to look at the, uh, criteria on the rubric, to say, "Oh, ok, I need to do this", and then they do it [.... B]ut then it's hard to always tell if they're really understanding it.

Nathan found a comparable shortcoming in the benchmark assignment, noting that students were not prompted to identify an audience for their summary memo:

> I understand they're bachelor's students, but many times they're not writing to their audience or what they, you know, what they – well, actually, what the exercises fail to describe as the audience [.... W]ho are you writing this summary memo to? Are writing it to file, are you writing it to the CEO, are you writing it to a boss? Because it's a different – you know, that can be a different animal.

111

As earlier in the chapter, the benchmark assignment and its rubric sometimes represented a hindrance to effective SLOA for these faculty, especially when they impeded faculty ability to leverage their expertise and experience for learning assessments. Because they view effective SLOA as so important to the students' growth, faculty members used the assignment-description issue to further solidify their conviction to an individual scoring approach.

   Representing the university.   In the healthcare field, faculty practitioners unofficially endorse the skillset and competence of university graduates they helped prepare in the classroom. Put another way, alumni job performance reflects onto the university and its faculty indirectly. Skilled workers represent their school well whereas poor performance could hurt the university's reputation and potentially reduce opportunities for future cohorts. This is especially important in ground campus locations, where the faculty and campus leaders maintain ties to local community resources, and new graduates seek employment from a limited range of firms. Nicole crystallized these concerns during one norming session:

> See, one of the other things is that, what I tell my students: by this time in their
>
> education, they're getting ready to go out into the workforce. And I don't want
>
> them not able to meet the expectations of, say, a healthcare organization in Philly.
>
> Um, because what they do is going to reflect on all of our students.

Erica alluded to the same apprehension when she described her sometimes-harsh scoring decisions on the benchmark assignment rubric:

> Knowing that this course is at the end of their juncture, you know, some of the
>
> things I see, I'm like, "whoa, whoa, whoa – I mean, I shouldn't see. And that's

112

how I feel. So, it's unfair to me – you say, you got this degree from SWU, and I –

especially me being at the end of your juncture – and I didn't say, "Red flags"?

The dread underpinning these thoughts is that other faculty members may not hold

sufficiently high standards and will hurt the end-product of competent alumni

contributing to their field and lending credence to the university. Their obligation to this

cause also seems to supersede consistency in SLOA with the benchmark assignment

rubric.

In sum, faculty perceptions of SLOA and their collaboration with others in

norming exercises impacted the consistency of their rubric scoring. Norming

participation was a positive experience for faculty members, who strengthened their

understanding of SLOA and specific student learning outcomes within their academic

program through productive interactions with colleagues. At the same time, faculty

clearly prioritized being able to give more accurate and meaningful assessments of

student learning by favoring an individualized approach to rubric scoring rather than

standardized, university-wide scoring protocols.

CHAPTER 5

DISCUSSION

A few years ago I began an exploration in my workplace to increase the

consistency with which SWU faculty used common rubrics to assess student learning

outcomes. I am edified by the action research it spurred, I am humbled by the support

given me academically and professionally to carry out the work, and I am motivated to

continue innovating toward long-term solutions. This chapter expresses these sentiments,

culminating the dissertation with a summary of the findings and how they contribute to

relevant research literature and the theoretical underpinnings of the study. This

discussion also includes a description of limitations and validity of this study, some

implications and next steps for continuing action research cycles, and general lessons I

took away from the experience.

Summary of Findings

Analysis of the study's quantitative and qualitative data sparked insights

regarding the effect of rubric norming on faculty participants for SLOA scoring using a

common rubric. Inter-rater reliability among scorers remained low throughout the study;

resource constraints, such as the benchmark assignment and its rubric, along with unclear

guidance from leadership, appear to have impacted this consistency. Perceptions of

SLOA have improved, especially regarding how to use assessment data and what the

institution does to encourage SLOA data use. But participants' survey responses and

observed discussions also suggest that assessment was always regarded as important. The

study's first two research questions posited that participation in a rubric-norming

intervention would increase both rating consistency and SLOA perceptions among

114

faculty participants. These findings not only answer those two research questions, they also point to reasons why the level of scoring consistency was lower than expected and why certain areas of faculty perceptions grew when others did not.

Collaboration among faculty participants revealed easy agreement on most scoring-related topics, a willingness to learn from others' experiences, and sincere effort toward identifying shared assessment goals. Though faculty members engaged in critical reflection and remained open to changes in practice, they rarely challenged differing viewpoints during norming and treated inconsistent rubric scoring among their colleagues as expected if not acceptable. The findings here fully addressed research question #3 by exploring the characteristics of collaboration among norming participants and linking the characteristics to themes of effective SLOA and assessment culture.

From a faculty perspective, university-wide or program-level assessment efforts often suffer from uneven communication and leadership. As a result, faculty participants lack confidence in overarching SLOA processes even while valuing collaboration with colleagues during assessment training. Instead, faculty participants lean on professional and academic standards they value and worry less about doing SLOA consistently. Faculty comments synthesized into what I now call the 'gatekeeper' mentality, an attitude facilitated and strengthened by participants' dual role as faculty-practitioners. Participants mentor students academically while shepherding them toward the health care industry in which they have deep experience. In a sense, they are 'vouching' for students upon entering the field and thus take great care to help them develop the requisite career-related competencies. This answers research question #4 and helps craft a more

115

comprehensive narrative around how the faculty participants in the research study attempt both effectiveness and consistency in their assessment classroom scoring.

Contributions to Theoretical and Research Literature

  Norming and inter-rater reliability. In this study, rubric norming did not impact statistically significant increases in inter-rater reliability among faculty participants. Many education research studies have shown that training can improve raters' scoring consistency (Hanssen et al., 2014; Oakleaf, 2009; O'Connell et al., 2016; Saxton et al., 2012). Other scholars found increases but qualified them, contending that positive inter-rater reliability effects of norming may only occur with specific rater populations or under particular conditions, and that the scoring tool used may play a role (Lovorn & Rezaei, 2011; Reddy & Andrade, 2010; Turbow & Evener, 2015). My study contributes to this ongoing conversation by providing a specific context in which a particular kind of rater training could not improve inter-rater reliability among faculty scorers.

  Moreover, I assert that specific contextual factors contributed to the lack of agreement among faculty raters. The training intervention may have been too weakly constructed to garner increases in inter-rater reliability. Live norming was limited to two hours per session, and faculty raters were only asked to score four work samples in each set of independently scored papers. Many researchers' studies suggested longer assessment trainings for increasing reliability and additional scoring practice for faculty raters using a common rubric (Holmes & Oakleaf, 2013; O'Connell et al., 2016; Preusche et al., 2012; Turbow & Evener, 2015). For the purposes of rater training set-up, the current research cycle establishes a lower boundary for norming, in terms of time spent and work samples scored to begin garnering statistical increases inter-rater reliability.

116

Other scholarship reminds that poorly-constructed scoring tools may lead to confusion among raters and low inter-rater reliability (Hack, 2015; Jonsson & Svingby, 2007). While the faculty raters in my study crafted shared learning expectations together, they were also navigating the messiness of student performance on the benchmark assignment. The study thus offers qualitative evidence as to how rater participants dealt with the interaction of these effects in a particular research context.

SLOA as a socially-constructed, collaborative process. A more important output of the study, though, is its contribution to the body of social constructivist research viewing SLOA as a collaborative, messy enterprise (Bloxham, 2009). SLOA is a necessarily subjective exercise, requiring expert raters to render judgments that are at least somewhat based on personal experience and bias (Price et al., 2008; Sadler, 2009). Because of this, assessment rater training is often rendered ineffectual because scorers – especially faculty or other professionals who represent an external accountability for the learning outcome – tend to backslide away from standard processes in favor of holistic and often incoherent personal scoring strategies (Sadler, 2009). In my study's norming intervention, faculty raters each touted their own individualized scoring rationale in at least one part of the rubric. When they could not find effective ways to integrate their judgment into the standardized scoring process – or when they felt the learning expectations had not been well communicated or practiced among their colleagues – they eschewed these artifacts in favor of their own assessment scoring schema. The resulting unevenness in rubric-scoring was only exacerbated if in fact faculty participants employed conformity strategies to feign agreement or avoid confrontation with one another. Future research in this area should probe deeper, and the context in which this

117

initial exploration occurred contributes to the field, potentially informing other scholars in the field and their cycles of inquiry.

Within higher education, within an institution, or within a specific academic college or program, shared standards of learning must be developed, and all stakeholders must buy in and benchmark student learning against them. To this end, some scholars have pointed out institutions' need for standards-sharing and development (Price et al., 2008). Additionally, because student learning outcomes represent a socially-constructed standard, proper assessment rater training, by 'norming' or any other name, should allow for individual rubric interpretations while still reinforcing reliability among raters' scores (Bloxham, 2009; O'Connell et al., 2016; Sadler, 2005). In its simplest sense, my study contributes to the literature in this field as a context-specific example of how a collaborative rubric-norming workshop can facilitate this kind of learning-expectations development among faculty colleagues. Though greater scoring consistency among raters was not achieved, the findings make clear that collaboration spurred understanding of shared learning outcomes and critical reflection on how those insights affect individual scoring schema.

Positive faculty perceptions of assessment. Faculty participants' collaboration within the research study stemmed their collective disposition toward assessment was positive. As such, the attitudes embodied by participants in my study contribute thoughtfully to the ongoing conversation on the role of faculty in assessment. I earlier noted faculty resistance to institutionally mandated assessment was an oft-cited issue in higher education scholarship (Cain & Hutchings, 2015; Shavelson, 2010), but that other researchers characterized faculty perceptions of SLOA differently (Danley-Scott & Scott,

2014; Rickards et al., 2016). My own experience recruiting faculty members for assessment research, and the study population that resulted, lend credence to both perspectives. A couple direct responses to my initial research invitation mimicked the fatigue and wariness for assessment work that others have predicted (Haviland, 2009). The participants who opted in, on the other hand, indicate a positive disposition toward SLOA and a desire to up-skill in relevant areas, consistent with other scholars' findings (Haviland et al., 2010; Schoepp & Benson, 2016). Instead of general resistance or aversion to SLOA work, the results of my mixed-methods approach to faculty perceptions of SLOA showed situational or structural issues limit faculty participation in SLOA. Non-participants responding to the study invitation cited hectic schedules or concern that our institution could not effectively lead assessment work, not that assessment itself was a meritless enterprise. Other researchers have pointed to these obstacles being especially poignant for adjunct faculty, who may appear unwilling to do extra work but run up against restrictive environmental factors (Danley-Scott & Scott, 2014). My research, and the voice of the faculty participants, contributes to the multi-layered discussion of how best to involve faculty members in higher education assessment.

Development of assessment culture through collaboration. Despite low inter-rater reliability and uneasiness with university leadership for assessment work, this study's findings show faculty developing positive perceptions of SLOA and improving individual scoring confidence. The true catalyst for this is collaborative interactions among faculty colleagues sharing similar goals. As norming participants exchanged ideas and developed a breadth of perspectives, they enriched their understanding of student

119

learning outcomes. Scholarship on SLOA in higher education shows this type of collaboration driving positive assessment outcomes for faculty participants in several different contexts (Guetterman & Mitchell, 2016; Haviland et al., 2010; Kogan et al., 2015). Furthermore, research shows that as SLOA participants collaborate toward insight and then critically reflect on practical improvements, they can drive more effective assessment efforts and a more positive culture of assessment as well (Guetterman & Mitchell, 2016; Holzweiss et al., 2016; Rodgers et al., 2013; Schoepp & Benson, 2016). My particular research context supports this idea. Survey results show that faculty participants' appreciation for collaboration increased significantly after the norming intervention, along with other perceptions of assessment. Further, they indicated a willingness to transform classroom assessment practices to better adhere to the shared expectations developed with colleagues during the norming exercises.

Finally, the study contributes to culture-of-assessment scholarship as another research context in which Guetterman & Mitchell's faculty perceptions survey has been adapted. In my study, the survey tool encouraged participants to describe their knowledge of SLOA, attitudes toward assessment, and institutional leadership for SLOA work. Results of the survey before and after the norming intervention reinforced the survey tool authors' own findings that collaboration with colleagues in assessment improves perceptions of SLOA and may point toward elements of a healthier culture of assessment at an institution (Guetterman & Mitchell, 2016). Though the study could not generate statistically significant increases in inter-rater reliability, other positive effects of rubric norming are evident through the observed collaboration and perceptions of SLOA

explained by faculty participants. These findings and the context in which they occurred constitute a contribution to SLOA research literature.

Assessment culture within activity systems analysis. From the activity-theoretical lens, the institution's culture of assessment and faculty attempts at rubric norming can be understood as a complex and interdependent set of sociocultural factors. Thus, faculty participants at SWU collectively demonstrated that their online, virtual spaces for interactive work functioned as an activity system. The artifacts used for collaborative norming, combined with social and community influences, impacted faculty raters' ability to consistently score student assessments. This dynamic among faculty raters supports research of Yamagata-Lynch who has used activity systems analysis to reveal tensions among educators and obstacles impeding their shared goals in a variety of contexts (Yamagata-Lynch & Haudenschild, 2009; Yamagata-Lynch, 2010). Scholarship further demonstrates that online collaborative spaces can generate the same the same activity-theoretical dynamic (Ryder & Yamagata-Lynch, 2014). In the case of faculty assessment participants, online discussions or score negotiation can still generate rich examples of collaborative strategies and evolution of shared goals (Trace, Meier, & Janssen, 2016). The current study echoes this point, showing a robust activity system reflected in the virtual rubric-norming space for faculty participants.

Reflecting on the activity systems analysis, the mixed-methods approach of my study matters as well. Fortunately, my study design and methodology was influenced by scholars who warned of the limited insights that purely quantitative studies of rater trainings yielded (Wang, 2010; Weigle, 1998). The norming intervention could not garner increased inter-rater reliability, but by blending both quantitative and qualitative research

121

in the data analysis phase, I was able to reveal several other positive results. Activity systems analysis helped when interpreting the integrated strands of findings, fitting the complicated and intertwined effects of shared learning expectations, individual scoring approaches, and inconsistent leadership around the research context into a sociocultural framework. Thus, even though the in-study interactions of norming participants could not increase scoring consistency, participants solidified their identities as members of academic and professional communities, bolstered perceptions of SLOA, and communicated these assessment-specific values to one another in rich ways. This signifies an improving culture of assessment at SWU and a true benefit of faculty participation in norming.

Implications for Research, Policy, & Practice

Research supports the idea that well-crafted, collaborative rater trainings can, in the long-term, improve inter-rater reliability for faculty participants (Hanssen et al., 2014; O'Connell et al., 2016), but that SLOA will always represent a set of irreducible, socially-constructed standards sensitive to individual rater bias (Bloxham et al., 2016). Further, this study's findings suggest that exposure to colleagues' perspectives and collaboration toward shared rubric interpretations clearly generated the most value for faculty participants. The first step forward then should be a re-evaluation of university priorities for SLOA initiatives. Leaders atop the institution, as well as those from individual colleges and academic programs should re-define the goal of rubric norming as a university-wide vehicle for faculty assessment training and the purpose for SLOA overall at the institution. The process to generate such discussions would be lengthy, and it may determine the currently employed model for SLOA in fact suffices. It is also

122

possible that the current study does not carry sufficient weight to spur that type of high-level response. In any case, there are other implications for research, policy, and practice stemming from the current study's context and findings.

Iterations and variations on current designs. If the current model for program-level outcomes assessment persists at SWU, future research cycles can be steered by basic variations on the current study's design, in an attempt to gauge the most efficient method of norming faculty to increase perceptions of SLOA or inter-rater reliability for rubric scores. This could include manipulation of the order or structure of the intervention phases, or through introduction of another treatment or control group into the study. Such adjustments could be prompted by decisions from college leaders seeking more impact to inter-rater reliability, or in response to assessment scholars advocating for more control-group experimentation for testing assessment innovations (Kezar, 2013). A related strand of research may involve revising the format, content, or delivery of rubric-norming. Asynchronous norming holds promise as a mechanism for faculty interactions related to SLOA, and libraries of anchor papers or benchmark rubric scores have also been floated as potential avenues for growth. Each of these, or any other similar changes, would require substantial testing to determine the most effective methods. Finally, if the university or an individual academic program determine that faculty assessment trainings should become even more collaborative or interactive, with a lower emphasis on inter-rater reliability, this too would affect the components of the training and necessitate some testing of elements before scaling to the broader university-wide population.

Regardless of the form that SLOA work takes in the future, different faculty populations will need to be included. Other colleges and programs exist, of course, and

123

any particular sub-population of faculty members might represent a different mix of demographics within SWU. Other colleges' faculty members may exhibit different changes in inter-rater reliability or demonstrate more/less growth in their perception of SLOA that shown here. The effects of collaboration among these populations too may differ from what occurred in this specific context too, so exploration beyond the current college or program, or the related professional area, is warranted.

Communication of Standard Practices. Although some areas of faculty perceptions for assessment improved after the rubric-norming intervention, participants clearly expressed a lack of real guidance for assessment scoring practice. Explicit rules for using the auto-scoring rubric tool are not well-communicated, rubric-language issues persist in many areas, and faculty generally perceive a laissez-faire attitude from leadership toward assessment. Participant Erica memorably said during a norming session, "Nothing is black and white." She was referring to the cloudiness of assessment protocol for faculty doing SLOA with benchmark rubrics, and how this made her care less about scoring reliability with her faculty colleagues. If SWU can sponsor more effective communication to its faculty on these policies – and clarify too where they still wanted to encourage individualized approaches to teaching and assessment work – it would represent a clear, quick win for the institution.

Two specific areas exist where such direction is needed. First, assessment leaders must definitively state the university policy in the ongoing 'scoring versus grading' debate among our faculty and staff. The dual-purpose rubrics and online rubric scoring tool allow faculty to overwrite the rubric-calculated score for students, allowing faculty to award a different point total for the course-assignment grade. This feature affords the

124

faculty greater freedom to account for course-specific elements that the grade should encompass, or credit the student for something outside the specific scope the assignment guidelines. As stated previously, not all faculty feel comfortable using this feature, so the university needs to make clear how the tool should be used and when overwriting of scores should take place. Second, university leaders have determined that accurate, consistent scoring with common rubrics is fostered by a deliberate, line-by-line, point-by-point approach to the tool. Faculty should be reminded of this before benchmark assignments begin in a given course week, and allowed to practice the skill to make it faster and easier. If faculty revert to sloppy, holistic judgments, the rubric's full potential goes untapped, and scoring consistency among raters suffers. Worse, shared understanding of learning expectations suffers, due to faculty reliance on inexact rubric-language interpretations. These two communication improvements will bolster scoring consistency for benchmark assignment rubrics, and faculty may also perceive university assessment leadership differently with clearer, more assertive communication in these areas.

Faculty collaboration, conformity, and agreement. When faculty participants traded thoughts on rubric scores, differences of opinion rarely generated direct debate or challenges regarding one's interpretation of the rubric. This phenomenon likely contributed to the low inter-rater reliability among faculty participants; thus how and why it occurred merits further exploration. It may be explained by sociocultural influences: faculty members may think it undesirable to create confrontation with colleagues they do not know well. They may seek to conform by instead welcoming another's perspective despite not truly valuing the comment. Or, they may truly have elaborated on their own

125

understanding of a specific student learning outcome, and they want to validate their colleague's point of view. In addition to possible conformity mechanisms at work in faculty dialogue, the social-constructivist paradigm posits that rubric-scoring discussions are rich with individual biases and personal assessment approaches as described throughout the dissertation. Qualitative analysis of these dialogues would do well to discern the sociocultural elements present, and how these affect scoring accuracy and scoring consistency for SWU faculty raters.

Qualitative focus on understanding SWU faculty members. Beyond further rubric-norming observations to investigate conformity, agreement, and the nuances of faculty collaboration for assessment, other areas of faculty practice merit potential research. I take away from this study a profound appreciation for the faculty participants' perspective for what assessment does for our students, and how it works best in the classroom. I anticipate sharing this appreciation with other SWU leaders, hopefully growing an appetite for more research into the faculty perspective on work at our institution. Different qualitative methodologies offer other research lenses and deeper, varied insights as to how they perceive their work and the behaviors underpinning instruction and assessment. I could concentrate an ethnographic inquiry on a single faculty member. Using both interviews and observations, the study could focus on the meaning they create professionally when interacting with their students, or a closely related topic. Alternately, university leadership may prefers to explore multiple faculty perspectives at once. If so, I could define a sub-population of interest based on purposeful sampling criteria and conduct a set of interviews or focus groups. This would allow for a diversity of opinions on a given topic, related to assessment or some other aspect of

126

classroom instruction. Finally, a sonic analysis of norming among faculty members holds potential. In this brand of study, I would illustrate what norming actually 'sounds' like, exploring and/or manipulating the audio produced by observed norming sessions among faculty members. This is a relatively non-traditional method and may also require additional software or technological resources. Such a study may potentially derive meaning and develop themes from the authentic sounds produced by participants. The presentation quality and novelty of such a study may present issues, but they would also guarantee a captive audience for sharing results afterward too. In all cases, the additional qualitative emphasis on faculty data-collection promises more insight as to what drives the teaching and learning of our SWU classrooms.

Validity and Limitations of Research Study

This mixed methods research study tested a training intervention to improve scoring consistency among faculty members. Some contextual factors were controlled; only faculty members instructing a specific course were invited, and the norming intervention focused on their use of one specific benchmark assignment rubric. I attempted to randomize others to minimize their potential effect on study results; for this reason, I did not invite specific faculty members within the list of those approved to teach HCS/475, and I chose student work samples – for practice scoring – without attempting to gauge the students' grade or writing skill level. Despite this approach, these variables may still have introduced non-random variance and unintended noise into the study. The mix of student work samples may not have been representative of the greater skill continuum among our student population. Similarly, the faculty who agreed to participate may differ from their absent colleagues in non-random ways, feeling more

127

positively about the university or engaging more readily with assessment initiatives. These unknown elements confound the results and may not reflect SWU's actual culture of assessment.

Other factors limited deeper analysis of the quantitative data. Practical considerations around budget and time restricted data collection. Having more raters or paying participants to score more student work samples would have garnered more data points. Fewer ratings on which to measure consistency impacted the potential significance power of the findings and increased the error of the calculated correlation coefficients. Small quantities of data succumb more easily to the random effects of specific raters' biases or the quality of a student work sample.

The qualitative data would have benefitted from additional coders; their help in data analysis would have provided more security against confirmation bias. Due to unforeseen scheduling issues, however, a second coder could not be retained for this round of action research. I combatted this potential reliability issue by often recording personal bias that arose within my research memos. This helped me to acknowledge and then distance myself from such influences. More importantly, as detailed in the third chapter, 'polyangulation' of multiple data sources safeguarded the validity of the qualitative data coding and member-checking allowed my faculty research participants to validate the accuracy of my analysis and interpretation. These methods, along with a careful, iterative approach to the data analysis, assisted me in protecting the study against unwitting and unwanted bias.

Lessons Learned

Perhaps the greatest insight I will take away from this study is an openness to faculty feedback. The university promotes the idea of consistency in rubric scoring as an aim for certain assessment training; some faculty members helped create the trainings and their goals. For my study, I assumed that all faculty participants would share this goal of inter-rater reliability when scoring with benchmark assignment rubrics, reinforced by norming exercises that developed common interpretations of rubric dimensions and learning expectations. Faculty members didn't quite see it that way, however. Their feedback throughout the study demonstrated a priority for effective SLOA in the classroom rather than 'consistent' SLOA. This distinction, focused on helping individual students get meaningful feedback on their attainment of learning outcomes, was not insignificant. Rather than assuming faculty motivation or goals, in any future iterations of this work, I will begin assessment workshops with a deliberate discussion of the workshop's aims, and the goals of SLOA in general. The richness of their in-session dialogues created by faculty participants reminds me also that the classroom faculty and the learning they facilitate should be regarded with appropriate respect.

On a similar note, the inherent value of a mixed-methods approach stands as another important lesson gleaned from the current research effort. I approached my problem of practice as a measurement issue initially, attempting to increase inter-rater reliability to bolster the student-learning inferences made with rubric data. It matters whether statistical consistency increased after rubric-norming workshops, but by itself, it constitutes an incomplete and narrow understanding of assessment training. Analyzing the interactions of my faculty participants, I was able to more fully explain how norming

129

affected study participants, and why it matters for SWU's assessment efforts more broadly. Mixed methodology fostered a comprehensive view of the problem – an appreciation that understanding our assessment culture at SWU was not a simple, cut-and-dry matter. Further, it has aided my ability to share findings and interpretations with multiple levels of stakeholders. When discussing the study, I concentrate on the richness of the data by playing authentic audio from norming interactions among faculty colleagues so that qualitative themes emerge from the real voices of faculty participants. It is a powerful vehicle; through it, I hope to continue improving our institution's culture of assessment alongside more effective and consistent student learning outcomes assessment.

REFERENCES

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Association of American Colleges & Universities. (2017).*On solid ground: VALUE report 2017*. Washington, D.C: AAC&U. Retrieved online from www.aacu.org.

Astin, A. W. (1991). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education.* New York: Macmillan.

Banta, T. W. & Palomba, C. A. (2014). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. San Fransisco, CA: Jossey-Bass.

Blackler, F. Cultural-historical activity theory and organization studies. In A. Sannino, H. Daniels, K. D. Gutierrez (Eds.), *Learning and expanding with activity theory.* Retrieved online from http://www.myilibrary.com?ID=239356

Bloxham, S. (2009). Marking and moderation in the UK: False assumptions and wasted resources. *Assessment & Evaluation in Higher Education, 34*(2), 209-220. doi: 10.1080/02602930801955978

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. doi: 10.1080/02602938.2015.1024607

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101. doi:10.1191/1478088706qp063oa

Bresciani, M. J., Oakleaf, M., Kolkhorst, F., Nebeker, C., Barlow, J., Duncan, K., & Hickmott, J. (2009). Examining design and inter-rater reliability of a rubric measuring research quality across multiple disciplines. *Practical Assessment, Research & Evaluation, 14*(12), 1-7. Retrieved online from http://pareonline.net/getvn.asp?v=14&n=12

Bresciani, M. J. (2011). Challenges in the implementation of outcomes-based assessment program review in a California community college district. *Community College Journal of Research and Practice*, *35*(11), 855-876. doi:10.1080/10668920802289877

Brookhart, S. M. & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, *67*(3), 343-368. doi:10.1080/00131911.2014.929565

Bullough, R. V., Jr. (2010). Proceed with caution: Interactive rules and teacher work sample scoring strategies, an ethnomethodological study. *Teachers College Record*, *112*(3), 775-810. Retrieved from http://www.tcrecord.org.

Cain, T. R., & Hutchings, P. (2015). Faculty and students: Assessment at the intersection of teaching and learning. In G. D. Kuh, S. O. Ikenberry, N. A. Jankowski, T. R. Cain, P. T. Ewell, P. Hutchings, & J. Kinzie (Eds.), *Using evidence of student learning to improve higher education* (pp.95-116). San Francisco, CA: Jossey-Bass.

Charmaz, K. (2006). *Constructing grounded theory*. London: SAGE.

Cordero de Noriega, D. & Diiorio, S. (2006). Administrative alignment and accountability: Student learning as focus. In A. Driscoll & D. Cordero de Noriega (Eds.), *Taking ownership of accreditation: Assessment processes that promote institutional improvement and faculty engagement* (229-242). Sterling, VA: Stylus.

Cresswell, J. W. (2015). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research.* (5th ed.). Boston, MA: Pearson

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process.* Thousand Oaks, CA: SAGE.

Danley-Scott, J. & Scott, G. (2014). The other half: Non-tenure track faculty thoughts on student learning outcomes assessment. *Research & Practice in Assessment, 9*, 31-44. Retrieved online from http://www.rpajournal.com/dev/wp-content/uploads/2014/06/A2.pdf

Darling-Hammond, L., Wilhoit, G., & Pettinger, L. (2014). Accountability for college and career readiness: Developing a new paradigm. *Education Policy Analysis Archives*, *22*(86). doi: http://dx.doi.org/10.14507/epaa.v22n86.2014

Dawson, P. (2015). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, *2015*. doi:10.1080/02602938.2015.1111294

Deneen, C. & Boud, D. (2014). Patterns of resistance in managing assessment change. *Assessment & Evaluation in Higher Education*, *39*(5), 577-591. doi:10.1080/02602938.2013.859654

132

Fairbanks, C. M., Duff, G. G. Faircloth, B. S., He, Y., Levin, B., Rohr, J., & Stein, C. (2010). Beyond knowledge: Exploring why some teachers are more thoughtfully adaptive than others. *Journal of Teacher Education*, *61*(1-2), 161-171. doi:10.1177/0022487109347874

Edwards, A. (2005). Relational agency: Learning to be a resourceful practitioner. International Journal of *Educational Research*, *43*, 168-182. doi:10.1016/j.ijer.2006.06.010

Edwards, A. (2007). Relational agency in professional practice: A CHAT analysis. *Actio: An International Journal of Human Activity Theory, 1*, 1-17. Retrieved online from http://www.chat.kansai-u.ac.jp/publications/actio/pdf/1_Edwards.pdf

Edwards, A. (2009). From the systemic to the relational: Relational agency and activity theory. In A. Sannino, H. Daniels, K. D. Gutierrez (Eds.), *Learning and expanding with activity theory.* Retrieved online from http://www.myilibrary.com?ID=239356

Edwards, A. (2011). Building common knowledge at the boundaries between professional practices: Relational agency and relational expertise in systems of distributed expertise. *International Journal of Educational Research*, *50*, 33-39. doi:10.1016/j.ijer.2011.04.007

Eldridge, D. (2016). Reliability and Validity of Instruments/Rubrics.[Sponsored conference presentation]. In *Program Impact: CAEPCon 2016 Fall Conference*. Washington, DC: Council for the Accreditation of Educator Preparation.

Eliam, G. (2003). The philosophical foundations of Aleksandr R. Luria's neuropsychology. *Science in Context*, *16*(4), 551-577. doi: 10.1017/S0269889703000966

Engeström, Y. (2001). Expansive learning at work: Toward an activity theory reconceptualization. *Journal of Education and Work*, *14*(1), 133-156. doi:10.1080/13639080020028747

Engeström, Y. & Hannele, K. (2007). Studies of expansive learning: Foundations, findings, and future challenges. *Journal of Workplace Learning, 6*(19), 336-342. Retrieved online from ProQuest *ebrary*.

Engeström, Y. & Sannino, A. (2010). From workplace learning to inter-organizational learning and back: The contribution of activity theory. *Educational Research Review*, *5*, 1-24. doi:10.1016/j.edurev.2009.12.002

133

Farquharson, K. (2013). Regulating sociology: Threshold learning outcomes and institutional isomorphism. *Journal of Sociology*, *49*(4), 486-500. doi:10.1177/1440783313504060

Friend, R., Rafferty, Y., & Bramel, D. (1990). A puzzling misinterpretation of the Asch 'conformity' study. *European Journal of Social Psychology, 20*(1), 29-44. doi: 10.1002/ejsp.2420200104

Foot, K. A. (2014) Cultural-historical activity theory: Exploring a theory to inform practice and research. *Journal of Human Behavior in the Social Environment*, *24*(3), 329-347. doi:10.1080/10911359.2013.831011

Fullan, M., Rincón-Gallardo, S., & Hargreaves, A. (2015). Professional capital as accountability. *Education Policy Analysis Archives, 23*(15). doi:10.14507/epaa.v23.1998

Fuller, M. B. (2012). Realizing higher education's humanizing potential: Assessment as a dialogical act. In G. Perreault & L. Zellner (Eds.), *Social justice, competition, and quality: 21st century leadership challenges, the 2012 yearbook of the NCPEA* (145-156). Retrieved online from http://www.ncpeapublications.org

Fuller, M. B. & Skidmore, S. T. (2014). An exploration of factors influencing institutional cultures of assessment. *International Journal of Education Research*, *65*, 9-21. doi: 10.1016/j.ijer.2014.01.001

Fuller, M. B., Skidmore, S. T., Bustamonte, R. M. & Holzweiss, P. C. (2016). Empirically examining higher education cultures of assessment. *The Review of Higher Education*, *39*(3), 395-429. doi:10.1353/rhe.2016.0022

Gade, S. (2015). Unpacking teacher–researcher collaboration with three theoretical frameworks: A case of expansive learning activity? *Cultural Studies of Science Education*, *10*, 603-619. doi: 10.1007/s11422-014-9619-7

Germaine, R. & Rubel Spencer, L. (2016). Faculty perceptions of a seven-year accreditation process. *Journal of Assessment and Institutional Effectiveness*, *6*(1), 67-98. Retrieved online from Project MUSE database.

Glăveanu, V. (2011). On culture and human development: Interview with Barbara Rogoff. *Europe's Journal of Psychology*, *7*(3), 408-418. doi: 10.5964/ejop.v7i3.141

Goodnough, K. (2016). Professional learning of K-6 teachers in science through collaborative action research: An activity theory analysis. *Journal of Science Teacher Education*, *27*(7), 757-767. doi: 10.1007/s10972-016-9485-0

Gorran Farkas, M. (2013). Building and sustaining a culture of assessment: Best practices for change leadership. *References Services Review*, *41*(1), 13-31. doi:10.1108/00907321311300857

Govaerts, M. J. B., Van de Wiel, M. W. J., Schuwirth, L. W. T., Van der Vleuten, C. P. M., & Muijtjens, A. M. M. (2013). Workplace-based assessment: raters' performance theories and constructs. *Advances in Health Sciences Education*, *18*(3), 375–396. http://doi.org/10.1007/s10459-012-9376-x

Graham, M., Milanowski, A., Miller, J. (2012). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings.* (Center for Educator Compensation Reform). Retrieved from ERIC database.

Greene, J. C. (2013). Reflections and ruminations. In D. M. Mertens & S. Hesse-Biber (Eds.), *Mixed methods and credibility of evidence in evaluation*. *New Directions for Evaluation, 138,* 109–119. doi: 10.1002/ev.20062

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*, 43-57. doi:10.1007/s10755-015-9330-y

Gutiérrez, K. D. & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, *32*(5), 19-25. doi: 10.3102/0013189X032005019

Hack, C. (2015). Analytical rubrics in higher education: A repository of empirical data. *British Journal of Educational Technology*, *49*(5), 924–927. doi:10.1111/bjet.12304

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23-34. Retrieved online from https://doaj.org/article/287241ad69104b63b4d2341173087f50

Handley, K. den Outer, B., & Price, M. (2013) Learning to mark: Exemplars, dialogue and participation in assessment communities. *Higher Education Research & Development*, *32*(6), 888-900. doi: 10.1080/07294360.2013.806438

Hanssen, E. E., Svensson, P. J., Strandberg, E. L.,Troein, M., & Beckman, A. (2014). Inter-rater reliability and agreement of rubrics for assessment of scientific writing. *Education*, *4*(1), 12-17. doi: 10.5923/j.edu.20140401.03.

Hargreaves, A. & Fullan, M. (2012). *Professional capital: Transforming teaching in every school.* NY: Columbia University Teachers College.

135

Haviland, D. (2009). Leading assessment: From faculty reluctance to engagement. *Academic Leadership*, *7*(1). Retrieved online from http://www.academicleadership.org/ideas_worth_merit/601.shtml

Haviland, D. (2014). Beyond compliance: Using organizational theory to unleash the potential of assessment. *Community College Journal of Research and Practice*, *38*, 755-765. doi: 10.1080/10668926.2012.711144

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*, *35*(4), 261-275. doi:10.1007/s10755-010-9140-1

Herr, K. & Anderson, G. L. (2015). *The action research dissertation* (*2nd ed*.). Thousand Oaks, CA: Sage.

Higher Learning Commission. (2015). *Adopted commission policy: Assumed Practices*. Retrieved from Higher Learning Commission website: http://www.hlcommission.org

Hodges, B. H. (2014). Rethinking conformity and imitation: Divergence, convergence, and social understanding. *Frontiers in Psychology, 5*. doi: 10.3389/fpsyg.2014.00726

Hodges, B. H. (2017). Carrying, caring, and conversing: Constraints on the emergence of cooperation, conformity, and language. *Interaction Studies*, *18*(1), 26-54. doi: 10.1075/is.18.1.02hod

Hodges, B. H., & Geyer, A. (2006). A Nonconformist Account of the Asch Experiments: Values, Pragmatics, and Moral Dilemmas. *Personality and Social Psychology Review, 10*(1), 2-19. doi: 10.1207/s15327957pspr1001_1

Hoffman-Kipp, P., Artiles, A. J., & Lopez-Torres, L. (2003). Beyond reflection: Teacher learning as praxis. *Theory into Practice*, *42*(3), 248-254. Retrieved online from http://www.jstor.org.ezproxy1.lib.asu.edu/stable/1477427

Holmboe, E. S., Sherbino, J., Long, D. M., Swing, S. R., & Frank, J. R. (2010). The role of assessment in competency-based medical education. *Medical Teacher*, *32*(8), 676-682. doi:10.3109/0142159X.2010.500704

Holmes, C. & Oakleaf, M. (2013). The official (and unofficial) rules for norming rubrics successfully. *The Journal of Academic Librarianship*, *39*(6), 599-602. doi:10.1016/j.acalib.2013.09.001

Holzweiss, P. C., Bustamonte, R. M., & Fuller, M. B. (2016). Institutional cultures of assessment: A qualitative study of administrator perspectives. *Journal of*

Assessment and Institutional Effectiveness, 6(1), 1-27. Retrieved online from https://muse.jhu.edu/article/623781

Hoyt, W., Kerns, M., & Appelbaum, M. I. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods, 4*(4), 403-424. doi: 10.1037/1082-989X.4.4.403

Hsieh, H. & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288. doi:10.1177/1049732305276687

Hutchings, P., Kinzie, J., & Kuh, G. D. (2015). Evidence of student learning: What counts and what matters for improvement. In G. D. Kuh, S. O. Ikenberry, N. Jankowski, T. R. Cain, P. T. Ewell, P. Hutchings, & J. Kinzie (Eds.), *Using evidence of student learning to improve higher education* (pp. 27-50). San Fransisco, CA: Jossey-Bass.

Jonson, J. L., Thompson, R. J., Guetterman, T. C. & Mitchell, N. (2016). The effect of informational characteristics and faculty knowledge and beliefs on the use of assessment. *Innovative Higher Education*, 42(1), 33-47. doi:10.1007/s10755-016-9366-7

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

Jonsson, A. & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity, and educational consequences. *Educational Research Review*, 2(2), 130-144. doi:10.1016/j.edurev.2007.05.002

Judd, T. P., Secolsky, C., & Allen, C. (2012, February 13). *Being confident about results from rubrics*. NILOA Guest Viewpoints Discussion Board: National Institute for Learning Outcomes Assessment, University of Illinois at Urbana-Champaign. Archived at https://blogs.illinois.edu/view/915/68373

Junor Clarke, P. A. & Fournillier, J. B. (2012). Action research, pedagogy, and activity theory: Tools facilitating two instructors' interpretations of the professional development of four preservice teachers. *Teaching and Teacher Education*, 28, 649-660. doi:10.1016/j.tate.2012.01.013

Kezar, A. (2013). Institutionalizing student outcomes assessment: The need for better research to inform practice. *Innovative Higher Education, 38*(3), 189-206. doi: 10.1007/s10755-012-9237-9

Kezar, A. & Maxey, D. (2014). *Student outcomes assessment among the new non-tenure-track faculty majority*. (Occasional Paper No. 21). Retrieved from National

Institute for Learning Outcomes Assessment website:
http://www.learningoutcomesassessment.org

Kinser, K. (2005). A profile of regionally accredited for-profit institutions of higher education. *New Directions for Higher Education*, *2005*(129), 69-83. doi:10.1002/he.173

Kinser, K. (2006). What Phoenix doesn't teach us about for-profit higher education. *Change: The magazine of higher learning, 38*(4), 24-29. doi:10.3200/CHNG.38.4.24-29

Kinser, K. (2007). Dimensions of corporate ownership in for-profit higher education. *The Review of Higher Education, 30*(3), 217-245. doi:10.1353/rhe.2007.0009

Knight, P., Tait, J., & Yorke, M. (2006). The professional learning of teachers in higher education. *Studies in Higher Education*, *31*(3), 319–339. doi: 10.1080/03075070600680786

Kogan, J. R., Conforti, L. N., Bernabeo, E., Iobst, W., & Holmboe, E. (2015). How faculty members experience workplace-based assessment rater training: a qualitative study. *Medical Education*, *49*(7), 692-708. doi:10.1111/medu.12733

Kuh, G. D., Jankowski, N., Ikenberry, S. O., & Kinzie, J. (2014). *Knowing what students know and can do: The current state of student learning outcomes assessment in U.S. colleges and universities.* Retrieved from National Institute for Learning Outcomes Assessment website: http://www.learningoutcomesassessment.org

Lakos, A. & Phipps, S. E. (2004). Creating a culture of assessment: A catalyst for organizational change. *portal: Libraries and the Academy*, *4*(3), 345-361. Retrieved online from the Project MUSE database.

Lechuga, V. M. (2008). Assessment, knowledge, and customer service: Contextualizing faculty work at for-profit colleges and universities. *The Review of Higher Education*, *31*(3), 287-303. doi: https://doi.org/10.1353/rhe.2008.0004

Lechuga, V. M. (2010). Who are they? And what do they do? In G. C. Hentschke, M. V. Lechuga, & W. G. Tierney (Eds.), *For-profit colleges and universities: Their markets, regulation, performance, and place in higher education* (pp. 46-70). Sterling, VA: Stylus. Retrieved from http://www.ebrary.com

Leont'ev, A. N. (1978). A*ctivity, consciousness, and personality.* (M. J. Hall, Trans.). Englewood Cliffs, NJ: Prentice-Hall.

138

Levy, P., & Williams, J. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management*, *30*(6), 881-905. doi:10.1016/j.jm.2004.06.005

Li, J. & Barnard, R. (2011). Academic tutors' beliefs about and practices of giving feedback on students' written assignments: A New Zealand case study. *Assessing Writing*, *16*, 137-148. doi:10.1016/j.asw.2011.02.004

Lovorn, M. G. & Rezaei, A. R. (2011). Assessing the assessment: Rubrics training for pre-service and new in-service teachers. *Practical Assessment, Research, & Evaluation*, *16*(16). Retrieved from http://pareonline.net/getvn.asp?v=16&n=16

Lyons, R. E. (2007). Deepening our understanding of adjunct faculty. In R. E. Lyons (Ed.), *Best practices for supporting adjunct faculty*. (pp.1-12). Bolton, MA: Anker Publishing.

Maki, P. (2010). *Assessing for learning: Building a sustainable commitment across the institution* (2nd ed.). Sterling, VA: Stylus.

McClellan, E. F. (2016). What a long strange trip it's been: Three decades of outcomes assessment in higher education. *PS: Political Science & Politics*, *49*(1), 88-92. doi: 10.1017/S1049096515001298

Mertler, C. A. (2014). *Action research: Improving schools and empowering educators* (4th ed.). Thousand Oaks, CA: SAGE.

Miller, R. & Leskes, A. (2005). *Levels of assessment: From the student to the institution.* (Greater Expectations Series). Retrieved from Association of American Colleges & Universities website: http://www.aacu.org

Mills, G. (2014). Dialogue in joint activity: Complementarity, convergence and conventionalization. *New Ideas in Psychology, 32*, 158-173. Retrieved online from http://dx.doi.org/10.1016/j.newideapsych.2013.03.006

Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: The Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 362*(1480), 639-648. doi: 10.1098/rstb.2006.2000

Ndoye, A. & Parker, M. A. (2010). Creating and sustaining a culture of assessment. *Planning for Higher Education*, *38*(2), 28-39. Retrieved online from http://search.ebscohost.com.ezproxy1.lib.asu.edu/login.aspx?direct=true&db=eft&AN=508123250&site=ehost-live

O'Connell, B., De Lange, P., Freeman, M., Hancock, P., Abraham, A., Howieson, B., & Watty, K. (2016). Does calibration reduce variability in the assessment of

139

accounting learning outcomes? *Assessment & Evaluation in Higher Education, 41*(3), 331-349. doi: 10.1080/02602938.2015.1008398

Oakleaf, M. (2009). Rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal of the American Society for Information Science and Technology*, *60*(5), 969-983. doi:10.1002/asi.21030

Paavola, S., Lipponen, L. & Hakkarainen, K. (2004). Models of innovative knowledge communities and three metaphors for learning. Review of Educational Research, 74(4), 557-576. Retrieved online from http://www.jstor.org/stable/3515981

Patton, K., Parker, M., & Tannehill, D. (2015). Helping teachers help themselves: Professional development that makes a difference. *NASSP Bulletin*, *99*(1), 26-42. doi:10.1177/0192636515576040

Pietarinen, J., Pyhältö, K., & Soini, T. (2016). Teachers professional agency – a relational approach to teacher learning. *Learning: Research and Practice*, *2*(2), 112-129. doi:10.1080/23735082.2016.1181196

Plano-Clark, V. L. & Cresswell, J. W. (2015). *Understanding research: A consumer's guide* (2nd ed.). Boston, MA: Pearson.

Preusche, I., Schmidts, M., & Wagner-Menghin, M. (2012). Twelve tips for designing and implementing a structured rater training in OSCEs. *Medical Teacher*, *34*, 368-372. doi: 10.3109/0142159X.2012.652705

Price, M., O'Donovan, B., Rust, C., & Carroll, J. (2008). "Assessment standards: A manifesto for change." *Brookes E-Journal of Learning and Teaching*, *2*(3). Retrieved online from http://bejlt.brookes.ac.uk/paper/assessment_standards_a_manifesto_for_change-2/

Reddy, Y. M. (2011). Design and development of rubrics to improve assessment outcomes: A pilot study in a Master's level business program in India. *Quality Assurance in Education, 19*(1), 84-104. doi:10.1108/09684881111107771

Reddy, Y. M. & Andrade, H. (2010). A review of rubric use in higher education. Assessment & Evaluation in Higher Education, 35(4), 435-448. doi:10.1080/02602930902862859

Rezaei, A. R. & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. Assessing Writing, 15, 18-39. doi:10.1016/j.asw.2010.01.003

Richardson, R. (2007). The associate program: A foundation for professional growth in adjunct faculty. In R. E. Lyons (Ed.), *Best practices for supporting adjunct faculty*. (pp.132-142). Bolton, MA: Anker Publishing.

140

Rickards, W H., Abromeit, J., Mentkowski, M., & Mernitz, H. (2016). Engaging faculty in an evaluative conversation. In W. H. Rickards & M. Stitt-Bergh (Eds.), *Evaluating student learning in higher education: Beyond the public rhetoric. New Directions for Evaluation*, *151*, 53–68. doi: 10.1002/ev.20198

Rodgers, M., Grays, M. P., Fulcher, K. H., & Jurich, D. P. (2013). Improving academic program assessment: A mixed-methods study. *Innovative Higher Education*, *38*, 383-395. doi:10.1007/s10755-012-9245-9

Rogoff, B. (1995). Observing sociocultural activity on three planes: Participatory appropriation, guided participation, and apprenticeship. In J. V. Wertsch, P. del Río, & A. Alvarez (Eds.), *Sociocultural studies of mind* (139-164). New York: Cambridge University Press.

Royal, K. (2011). Understanding reliability in higher education student learning outcomes assessment. *Quality Approaches in Higher Education*, *2*(2), 8-15. Retrieved from http://www.asq.org

Ryder, L. & Yamagata-Lynch, L. (2014). Understanding tensions: Activity systems analysis of Transpacific collaboration. *CALICO Journal*, *31*(2), 201-220. doi: 10.11139/cj.31.2.201-220.

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, *30*(2), 175-194. doi: 10.1080/0260293042000264262

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, *34*(2), 159-179. doi: 10.1080/02602930801956059

Sadler, D. R. (2013). Assuring academic achievement standards: From moderation to calibration. *Assessment in Education: Principles, Policy, & Practice*, *20*(1), 5-19. doi: 10.1080/0969594X.2012.714742

Sannino, A, Daniels, H., & Gutierrez, K. D. (2009). Activity theory between historical engagement and future-making practice. In A. Sannino, H. Daniels, K. D. Gutierrez (Eds.), *Learning and expanding with activity theory.* Retrieved online from http://www.myilibrary.com?ID=239356

Saxton, E., Belanger, S., & Becker, W. (2012). The Critical Thinking Analytic Rubric (CTAR): Investigating intra-rater and inter-rater reliability of a scoring mechanism for critical thinking performance assessments. *Assessing Writing*, *17*, 251-270. doi:10.1016/j.asw.2012.07.002

Schein, E. H. (2010). *Organizational culture and leadership (4th ed.).* San Fransisco, CA: Jossey-Bass.

Schilling, K. M. & Schilling, K. L. (1998*). Proclaiming and sustaining excellence: Assessment as a faculty role.* ASHE-ERIC Higher Education Report, v26, n3. Washington, D.C.: The George Washington University, Graduate School of Education and Human Development.

Schneider, B., Ehrhart, M. G. & Macey, W. H. (2013) Organizational climate and culture. *The Annual Review of Psychology*, *64*, 361-388. doi:10.1146/annurev-psych-113011-143809

Schoepp, K. & Benson, S. (2016). Meta-assessment: Assessing the learning outcomes assessment program. *Innovative Higher Education*, *41*, 287–301. doi:10.1007/s10755-015-9347-2

Schuetz, P. (2002). Instructional practices of part-time and full-time faculty. *New Directions for Community Colleges*, *118*, 39-46. doi:10.1002/cc.62

Scott, G. & Danley-Scott, J. (2015). Two loops that need closing: Contingent faculty perceptions of outcomes assessment. The Journal of General Education, 64(1), 30-55. doi: https://doi.org/10.1353/jge.2015.0004

Secolsky, C., Sng, C., Wentland, E., & Smith III, D. L. (2016). Assessment of student learning outcomes: Evaluator and client communication. In W. H. Rickards & M. Stitt-Bergh (Eds.), *Evaluating student learning in higher education: Beyond the public rhetoric. New Directions for Evaluation, 151,* 97–107. doi:10.1002/ev.20205

Selingo, J. J. (2016). *2026, the decade ahead: The seismic shifts transforming the future of higher education.* Washington, DC: The Chronicle of Higher Education.

Shanahan, M. (2012). Activity theory. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.), *Encyclopedia of case study research*. Thousand Oaks, CA: SAGE. doi: 10.4135/9781412957397

Shavelson, R. J. (2010). *Measuring college responsibly: Accountability in a new era.* Stanford, CA: Stanford University Press.d

Shulman, L. S. (2007). Counting and recounting: Assessment and the quest for accountability. *Change*, *39*(1), 20-25. Retrieved from: http://login.ezproxy1.lib.asu.edu/login?url=http://search.proquest.com.ezproxy1.lib.asu.edu/docview/208048672?accountid=4485

Southern Association of Colleges and Schools Commission on Colleges. (2012). *The principles of accreditation: Foundations for quality enhancement* (*5th ed*.). Retrieved from Southern Association of Colleges and Schools Commission on Colleges website: http://www.sacscoc.org

Stake, R. E. & Trumbull, D. J. (1982). Naturalistic generalizations. In M. Belok & N.
Haggerson (Eds.), *Review Journal of Philosophy & Social Science*, *7*(1&2).
Retrieved online from
http://education.illinois.edu/circe/Publications/CIRCE_Publications.html

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement
approaches to estimating interrater reliability. *Practical Assessment, Research, &
Evaluation*, *9*(4). Retrieved online from: http://pareonline.net/getvn.asp?v=9&n=4

Strand, P., Edgren, G., Borna, P., Lindgren, S., Wichmann-Hansen, G., & Stalmeijer, R.
E. (2015). Conceptions of how a learning or teaching curriculum, workplace
culture and agency of individuals shape medical student learning and supervisory
practices in the clinical workplace. *Advances in Health Science Education*, *20*,
531-557. doi: 10.1007/s10459-014-9546-0

Suskie, L. (2009). *Assessing student learning: A common sense guide (2nd ed.)*. San
Fransisco, CA: Jossey-Bass.

Thorndike, R. M. (2011). Fisher's Z transformation. In N. J. Salkind (Ed.), *Encyclopedia
of Measurement and Statistics*. Thousand Oaks, CA: SAGE. doi:
10.4135/9781412952644

Toma, J. D. (2008). A few reflections on accountability, affordability, and access
worldwide. *New Directions for Institutional Research*, *2008*, 93-97.
doi:10.1002/ir.281.

Trace, J., Meier, V., & Janssen, G. (2016). "I can see that": Developing shared rubric
category interpretations through score negotiation. *Assessing Writing*, *30*, 32-46.
doi:10.1016/j.asw.2016.08.001

Turbow, D. J. & Evener, J. (2016). Norming a VALUE rubric to assess graduate
information literacy skills. *Journal of the Medical Library Association*, *104*(3),
209-214. doi: 10.3163/1536-5050.104.3.005

U.S. Department of Education, National Center for Education Statistics. (2002). *1993
National Study of Postsecondary Faculty (NSOPF:93), Part-time Instructional
Faculty and Staff: Who They Are, What They Do, and What They Think,* NCES
2002–163, Washington, DC. Retrieved from:
http://nces.ed.gov/pubs2002/2002163.pdf

U.S. Department of Education. (2006, September). *A test of leadership: Charting the
future of U.S. higher education.* Retrieved from:
http://www2.ed.gov/about/bdscomm/list/hiedfuture/reports/final-report.pdf

143

U.S. Department of Education, National Center for Education Statistics. (2014). *Full-time student* [glossary]. Retrieved from: https://surveys.nces.ed.gov/ipeds/VisGlossaryAll.aspx

U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (2015, September). *Demographic and Enrollment Characteristics of Nontraditional Undergraduates: 2011-12.* NCES 2015-025, Washington, DC. Retrieved from: http://nces.ed.gov/pubs2015/2015025.pdf

Van Houtte, M. (2005). Climate or culture? A plea for conceptual clarity in school effectiveness research. *School Effectiveness and School Improvement*, *16*(1), 71-89. doi:10.1080/09243450500113977

Van Houtte, M. & Van Maele, D. (2011). The black box revelation: In search of conceptual clarity regarding climate and culture in school effectiveness research. *Oxford Review of Education*, *37*(4), 505-524. doi:10.1080/03054985.2011.595552

Waitoller, F. R. & Artiles, A. J. (2016). Teacher learning as curating: Becoming inclusive educators in school/ university partnerships. *Teaching and Teacher Education*, *59*, 360-371. doi: 10.1016/j.tate.2016.07.007

Walvoord, B. E. (2004). *Assessment clear and simple: A practical guide for institutions, departments, and general education*. San Fransisco, CA: Jossey-Bass.

Wang, B. (2010). On rater agreement and training. *English Language Teaching*, *3*(1), 108-112. Retrieved online from: http://search.proquest.com/docview/839755924?accountid=35812

Waugh, C. K. & Gronlund, N. E. (Eds.). (2013). Assessment of student achievement. (10th ed.). Upper Saddle River, NJ: Pearson.

Weigle, S. C. (1998). Using FACETS to model rater training effects. *Language Testing, 15*(2), 263-287. doi:10.1177/026553229801500205

Weigle, S. C. (1999). Investigating rater/prompt interactions in writing assessment: Quantitative and qualitative approaches. *Assessing Writing*, *6*(2), 145-178. doi: 10.1016/S1075-2935(00)00010-6

Weiner, W. F. (2009). Establishing a culture of assessment: Fifteen elements of assessment success. *Academe*, *95*(4), 28-32. Retrieved online from http://www.aaup.org/article/establishing-culture-assessment#.U7yrTrFnCSo.

Wertsch, J. V. (1995). The need for action in sociocultural research. In J. V. Wertsch, P. del Río, & A. Alvarez (Eds.), *Sociocultural studies of mind* (56-74). New York: Cambridge University Press.

Williams R.G., Klamen, D.A., & McCaghie, W.C. (2003). Cognitive, social and environmental sources of bias in clinical performance ratings. *Teaching and Learning in Medicine*, *15*(4), 270–292. doi: 10.1207/S15328015TLM1504_11.

Wiseman, C. S. (2012). Rater effects: Ego engagement in rater decision-making. *Assessing Writing*, *17*(3), 150-173. doi:10.1016/j.asw.2011.12.001

Woehr, D. J. & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. Journal of Occupational and Organizational Psychology, 67(3), 189-205. Retrieved online from EBSCO Host.

Wolfe, E.W., Matthews, S., & Vickers, D. (2010). The effectiveness and efficiency of distributed online, regional online, and regional face-to-face training for writing assessment raters. *Journal of Technology, Learning, and Assessment, 10*(1). Retrieved online from http://www.jtla.org

Wright, S. (2015). Relational agency from a teacher as researcher perspective. *Cultural Studies of Science Education*, *10*, 629-636. doi:10.1007/s11422-015-9664-x

Yamagata-Lynch, L. C. (2010). *Activity systems analysis methods: Understanding complex learning environments.* Springer US. Retrieved online from http://www.myilibrary.com?ID=292716

Yamagata-Lynch, L. C. & Haudenschild, M. T. (2009). Using activity system analysis to identify inner contradictions in teacher professional development. *Teaching and Teacher Education*, *25*, 507-517. doi:10.1016/j.tate.2008.09.014

Zuber-Skerritt, O. (1992). *Action research in higher education*. London: Kogan Page.

APPENDIX A

RECRUITMENT AND CONSENT FORM

146

Dear faculty,

My name is Nick Williams. I am a doctoral student in the Mary Lou Fulton Teachers College (MLFTC) at Arizona State University. I am working under the direction of Dr. Daniel Liou, a faculty member in MLFTC. I am conducting a research study on student learning outcomes assessment at UOPX, specifically through the perspective of faculty members engaged in norming, or calibration, training for use of benchmark-assignment rubrics in the classroom. I aim to examine the effectiveness of 'norming' training for increasing inter-rater reliability among faculty using a common rubric, and how training can affect faculty perceptions of assessment and how it might help foster a 'culture of assessment'. This study is being done toward partial completion of a doctoral program at Arizona State University. It is also being done in collaboration with the School of Health Services Administration at UOPX, of which all of you are active faculty members.

I am asking for your help, which will involve your participation in an online survey about your knowledge, attitudes, and beliefs about leading innovations for improved reliability of our assessment data. I anticipate the faculty survey will take about 15 minutes, and the scoring of ten sample papers using the benchmark-assignment will take approximately an hour and a half. These data-collection measures will be administered a second time in late November. You will also be asked to participate in training workshops, which will last two hours for a live, virtual session with colleagues, and follow-up engagement on message boards with colleagues, at least one time, for at least 30 minutes. Some participants will be asked to perform these exercises in October, and others will do so in January.

You have the right not to answer any question on the survey or in a training session, as well as the right to stop participation at any time. Your participation in this study is voluntary. If you choose not to participate or to withdraw from the
study at any time, there will be no penalty whatsoever.

For the data-collection measures and for the training components, you will be compensated for your time by the School of Health Services Administration at the rate of $25/hour, for up to a total of 6 hours.

In addition to compensation, the benefits to participation in the research include increased engagement and networking with colleagues, and the chance to reflect on your professional gains in the area of assessment. Survey results and successful components of the study may also inform future iterations of
the project. Thus, there is potential to enhance the experiences that are provided to our faculty and other campus/college leaders during our assessment processes. There are no foreseeable risks or discomforts to your participation.

Your responses will be confidential. Results of this study may be used in reports, presentations, or
publications but your name will not be known. I will also be audio-recording the training

147

workshops. The training will not be recorded without your permission. Please let me know if you do <u>not</u> want to the training to be recorded; you can also change your mind after the session begins, and you will be allowed to discontinue your participation.

Please read the following consent statement. If you agree, please sign the and print your check the box indicating your informed consent, and click the "next" button to give consent and proceed to the rest of the survey.

Consent Statement: I agree to participate in the surveys, training, and scoring protocols being conducted. I understand the survey will take approximately 15 minutes to complete, the scoring may take up to 90 minutes, and live training components will last two hours.
I understand that neither my relationship with the college nor with my campus will be affected if I opt out of taking the survey or participate in the interview. I am at least 18 years of age.

If you have any questions concerning the research study, please contact Daniel Liou at dliou@asu.edu or (XXX) XXX-XXXX or Nick Williams at nick.williams@phoenix.edu or (602) 557-2382. If you have any questions about your rights as
a participant in this research, or if you feel you have been placed at risk, you can contact the Chair of the Human Subjects Institutional Review Board through the ASU Office of Research Integrity and Assurance at (480) 965-6788.

Thank you,

Nick Williams, Doctoral Student
Daniel Liou, Assistant Professor

By signing below, you are agreeing to be part of the study, and consenting to the above bolded statement.

Name: _____ (print name)

Signature: _____ (sign name)

Date: _____ (mm/dd/yyyy)

APPENDIX B

IRB APROVAL – ARIZONA STATE UNIVERSITY

149

EXEMPTION GRANTED

Daniel Dinn-You Liou
Division of Educational Leadership and Innovation - West -
dliou@asu.edu

Dear Daniel Dinn-You Liou:

On 11/16/2015 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Norming at Scale: Faculty Perceptions of Assessment Culture and Student Learning Outcomes Assessment
Investigator:	Daniel Dinn-You Liou
IRB ID:	STUDY00003441
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• Nick Williams IRB Rec Consent, Category: Consent Form; • Assessment Attitudes and Knowledge Survey.pdf, Category: Recruitment Materials; • Rubric Norming - Session Guide, Category: Recruitment Materials; • Rubric Norming - Facilitator Prompts.pdf, Category: Recruitment Materials; • Letter of Support, Category: Off-site authorizations (school permission, other IRB approvals, Tribal permission etc); • Rubric Norming - Scoring Rubric, Category: Recruitment Materials; • IRB Protocol, Category: IRB Protocol;

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (2) Tests, surveys, interviews, or observation on 11/16/2015.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,
IRB Administrator
cc: Nicholas Williams

APPENDIX C

IRB RELIANCE AGREEMENT

University of Phoenix®

IRB - Reliance Agreement

The Officials signing below agree that the **University of Phoenix** may rely on the designated IRB **Arizona State University** for review and continuing oversight of its human subjects research described below: (check one)

( ) This agreement applies to all human subjects research covered by **Arizona State University** FWA.

(x) This agreement is limited to the following specific protocol(s):

Name of Research Project: **Norming at Scale: Faculty Perceptions of Assessment Culture and Student Learning Outcomes Assessment**

IRB/Protocol #: **Arizona State University, IRB # STUDY00003441**

Name of Principal Investigator: **Dr. Daniel Liou; Co-PI: Nick Williams**

Sponsor or Funding Agency: **NONE**   Award Number, if any: NONE

The review performed by the designated IRB, **Arizona State University**, will meet the human subject protection requirements of the University of Phoenix OHRP-approved FWA. The IRB at **Arizona State University** will follow written procedures for reporting its findings and actions to appropriate officials at the **University of Phoenix**. Relevant minutes of IRB meetings will be made available to the **University of Phoenix** upon request. **Arizona State University** remains responsible for ensuring compliance with the IRB's determinations and with the Terms of its OHRP-approved FWA. This document must be kept on file by both parties and provided to OHRP upon request.

v.0115

Signature of Signatory Official **University of Phoenix:**

University of Phoenix FWA #: 00004202

_____ *W.e.h☺* _____   Date: 8/9/2017

Print Full Name: Dr. William Beck

Institutional Title: IRB Chair

Signature of Signatory Official **Arizona State University:**

Arizona State University FWA #: 00009102

_____   Date: _____

Print Full Name: Debra Murphy

Institutional Title: Institutional Official

v.0115

153

APPENDIX D

ASSESSMENT KNOWLEDGE AND ATTITUDES SURVEY

154

Assessment Attitudes and Knowledge Survey

Q5 **Content knowledge** (1) How knowledgeable are you about the:

	Not at all (1)	Very Little (2)	Some what (3)	Quite a bit (4)	To a great extent (5)
a. Purpose of outcomes assessment? (1)	❍	❍	❍	❍	❍
b. Methods of outcomes assessment? (2)	❍	❍	❍	❍	❍
c. Institution's general education program? (3)	❍	❍	❍	❍	❍
d. General education learning outcome(s)? (4)	❍	❍	❍	❍	❍
e. Specific learning objective addressed in this project? (5)	❍	❍	❍	❍	❍

Q7 **Beliefs about assessment** (2) Indicate the extent to which you agree with the following statements.

	Not at all (1)	Very Little (2)	Somewhat (3)	Quite a bit (4)	To a great extent (5)
a. Assessment is primarily about improving student learning (1)	❍	❍	❍	❍	❍
b. Assessment is primarily about being accountable for student learning (2)	❍	❍	❍	❍	❍
c. I have yet to be convinced of the alleged benefits of outcomes assessment (7)	❍	❍	❍	❍	❍
d. Initiating a process for outcomes assessment would enhance the stature of our department/program/university (8)	❍	❍	❍	❍	❍
e. Outcomes assessment would pave the way for better programs for our students (6)	❍	❍	❍	❍	❍
f. It would be difficult to implement a procedure for outcomes assessment without seriously disrupting other activities (5)	❍	❍	❍	❍	❍

Q11 (3) Please describe how you feel about the following statements regarding student assessment at your institution.

	Disagree strongly (1)	Disagree somewhat (2)	Neither Agree nor Disagree (3)	Agree somewhat (4)	Agree strongly (5)
a. Students today are learning more due to an institutional focus on the assessment of student learning (1)	❍	❍	❍	❍	❍
b. Student assessment has improved the quality of education at this institution (2)	❍	❍	❍	❍	❍
c. Faculty use student assessment information to modify how or what they teach (3)	❍	❍	❍	❍	❍
d. Assessing students has resulted in the development of learning experiences that better meet diverse learning styles (4)	❍	❍	❍	❍	❍
e. Faculty update their in-class assessment techniques on a regular basis (5)	❍	❍	❍	❍	❍
f. The effectiveness of teaching is enhanced when faculty regularly engage in student assessment (6)	❍	❍	❍	❍	❍
g. Student assessment techniques accurately measure students learning (7)	❍	❍	❍	❍	❍

Q8 **Use of assessment findings** (4) In your opinion how useful is assessment of student learning in informing the following insights

	Not Useful (1)	Marginally Useful (2)	Somewhat Useful (3)	Useful (4)	Very Useful (5)
a. Improve program or practice (1)	❍	❍	❍	❍	❍
b. Influence thinking rather than action (2)	❍	❍	❍	❍	❍
c. Determine the overall worth or merit of a program (3)	❍	❍	❍	❍	❍
d. Mobilize support and legitimate a position (4)	❍	❍	❍	❍	❍

156

Q12 **University use of assessment findings** Colleges and universities increasingly use surveys and other measures (e.g., portfolios) to gather information about student educational experiences and learning. The following questions ask about your view of the University's assessment efforts.

Q13 (5) To what extent is the University involved in student assessment efforts?
- ◯ Very much (4)
- ◯ Quite a bit (3)
- ◯ Some (2)
- ◯ Very little (1)

Q14 (6) How effectively does the University disseminate the findings of its assessment efforts to faculty?
- ◯ Very effectively (5)
- ◯ Somewhat effectively (4)
- ◯ Neither effective nor ineffective (3)
- ◯ Somewhat ineffectively (2)
- ◯ Not at all effectively (1)

Q15 (7) In general, how useful to you are the findings from the University's assessment efforts?
- ◯ Very useful (5)
- ◯ Mostly useful (4)
- ◯ Neutral (3)
- ◯ Mostly not useful (2)
- ◯ Not at all useful (1)

Q16 (8) To what extent are results from the University's assessment efforts used to inform the following?

	Very much (1)	Quite a bit (2)	Some (3)	Very little (7)
a. Institutional activities aimed at improving teaching and learning (1)	◯	◯	◯	◯
b. Your department's activities aimed at improving teaching and learning (2)	◯	◯	◯	◯

Q17 (9) To what extent is evidence gathered by faculty members in their courses used to inform the following?

	Very much (1)	Quite a bit (2)	Some (3)	Very little (4)
a. Institutional activities aimed at improving teaching and learning (1)	◯	◯	◯	◯
b. Your department's activities aimed at improving teaching and learning (2)	◯	◯	◯	◯

157

Q18 (10) To what extent are faculty members at the University encouraged to do the following?

	Very much (1)	Quite a bit (2)	Some (3)	Very little (4)
a. Systematically collect information about the effectiveness of their teaching beyond end-of-term course evaluations (1)	○	○	○	○
b. Use assessment findings to inform changes made to their courses (2)	○	○	○	○
c. Publicly present (e.g., lectures or workshops) information about teaching or learning (3)	○	○	○	○
d. Publish on teaching and learning (4)	○	○	○	○
e. Collaborate with colleagues on improving teaching and learning (5)	○	○	○	○

Q19 (11) To what extent have you incorporated the following into your work?

	Very much (1)	Quite a bit (2)	Some (3)	Very little (4)
a. Systematically collecting information about the effectiveness of your teaching beyond end-of-term course evaluations (1)	○	○	○	○
b. Using assessment findings to inform changes made to your courses (2)	○	○	○	○
c. Publicly presenting (e.g., lectures or workshops) information about teaching or learning (3)	○	○	○	○
d. Publishing on teaching and learning (4)	○	○	○	○
e. Collaborating with colleagues on improving teaching and learning (5)	○	○	○	○

APPENDIX E

NORMING INTERVENTION BENCHMARK ASSIGNMENT AND SCORING

RUBRIC

www.manaraa.com

Rubrix (New Classroom)
Analytic Rubric Building Guidelines and Template

Course ID: HCS/475
Course Title: Leadership and Performance Development
Benchmark Assignment Title: Week 5, Summary Memo
Total number of points: 15

Resources: Problem Analysis Worksheet, Week Five Case Studies, and Summary Memo Guidelines

Review the Problem Analysis Worksheet.

Select one of the solutions you proposed in the Problem Analysis Worksheet.

Write a 350- to 700-word summary memo explaining why your solution will effectively resolve the conflict, how you propose to implement your solution, and your role as a leader to manage conflict and create an effective work environment.

Include the following in your summary:
- Summarize the problem and the solution you propose to implement.
- Analyze why you think the solution will be effective.
- Analyze what needs to be considered when implementing the proposed solution.
- Analyze the leadership style that best fits implementing the proposed solution.
- Analyze a leader's role in managing conflict.
    - What conflicts may arise from the problem or proposed solution, and what role should a leader take to manage those conflicts?
- Explain the leader's role in creating an effective work group when implementing the proposed solution.
- Summarize key points and next steps.

Include a reference page with your summary, and **cite** at least 2 references using APA guidelines.

Click the Assignment Files tab to submit your assignment.

PSLO: [Type in program SLO, as applicable, e.g. MAED/AET 1: 2-word descriptive title]	GE-SLO: ULG:	Dimensions or Assignment Criteria [Type in criterion title or summary phrase] Weight: [indicate weight of rubric line out of 100%]	Does Not Meet Expectations (1.00) 25%	Approaches Expectations (2.00) 50%	Meets Expectations (3.00) 75%	Exceeds Expectations (4.00) 100%
PSLO: BSHA PSLO 6: Management and Leadership	ULG: 2	Content: Problem and Solution - Introduction Weight: 12%	Did not summarize the problem or the proposed solution.	Accurately summarized the problem or the proposed solution, but not both.	Accurately summarized both the problem and the proposed solution.	Accurately summarized both the problem and the proposed solution. Provided a unique or creative summary that captured the reader's attention.
BSHA PSLO 6: Management and Leadership	ULG: 2	Content: Analyze Solution Weight: 12%	Did not identify or explain why the solution will be effective. May not have completed this element of the assignment.	Explained why the solution will be effective.	Analyzed why the solution will be effective.	Analyzed why the solution will be effective. Provided a unique perspective or depth of the topic.
BSHA PSLO 6: Management and Leadership	ULG: 2	Content: Analyze Solution Implementation Weight: 12%	Did not identify or explain what needs to be considered when implementing the proposed solution. May not have completed this element of the assignment.	Explained what needs to be considered when implementing the proposed solution.	Analyzed what needs to be considered when implementing the proposed solution.	Analyzed what needs to be considered when implementing the proposed solution. Provided a unique perspective or depth of the topic.
BSHA PSLO 6: Manageme	ULG: 2	Content:	Identified the leadership style that best fits	Explained the leadership style that best fits	Analyzed the leadership style that best fits	Analyzed the leadership style that best fits

161

nt and Leadership		**Analyze Leadership Style** *Weight: 12%*	implementing the proposed solution or did not complete this element of the assignment.	implementing the proposed solution.	implementing the proposed solution.	implementing the proposed solution. Provided a unique perspective or depth of the topic.
BSHA PSLO 6: Management and Leadership	**ULG: 2**	**Content:** **Analyze Leader's Role in Conflict** *Weight: 12%*	Identified a leader's role in managing the conflict or did not complete this element of the assignment.	Explained a leader's role in managing the conflict.	Analyzed a leader's role in managing conflict.	Analyzed a leader's role in managing conflict. Provided a unique perspective or depth of the topic.
BSHA PSLO 6: Management and Leadership	**ULG: 2**	**Content:** **Leader's Role in Effective Workgroups** *Weight: 12%*	Did not identify the leader's role in creating an effective work group or did not complete this element of the assignment.	Identified the leader's role in creating an effective work group when implementing the proposed solution.	Explained the leader's role in creating an effective work group when implementing the proposed solution.	Explained the leader's role in creating an effective work group when implementing the proposed solution. Provided a unique perspective or depth of the topic.
BSHA PSLO 6: Management and Leadership	**ULG: 2**	**Content:** **Key Points and Next Steps - Conclusion** *Weight: 12%*	Conclusion missing or did not summarize the key points and did not identify the next steps.	Conclusion summarized the key points or identified the next steps.	Conclusion summarized the key points and identified the next steps leaving the reader with a sense of closure.	Conclusion summarized the key points and identified the next steps drawing the reader to closure and resolution.
	ULG: 3 **GE**: 2	**Summary Memo Guidelines:** **Punctuation, Mechanics, Syntax, and** **Paragraph Quality and Fflow**	Excessive occurrences of writing errors detracted from the content. Language choices were unclear and did not exhibit a professional tone.	Frequent occurrences of writing errors detracted from the content. Language choices were commonplace and approached a	Occasional occurrences of writing errors did not detract from the content. Language choices were thoughtful and maintained a professional tone.	Rare occurrences of writing errors enhanced the content. Language choices were compelling and enhanced the professional tone.

162

		Weight: 10%	Paragraphs were lengthy and were not presented in a thoughtful order.	professional tone. Paragraphs were lengthy or were not presented in a thoughtful order.	Paragraphs were typically concise and presented in a thoughtful order.	Paragraphs were consistently concise and presented in a logical order.
	ULG: 4 **GE**: 6	**Information Utilization:** **Citations** *Weight: 6%*	Did not cite any peer-reviewed, scholarly, or similar references to support the assignment.	Cited one peer-reviewed, scholarly, or similar references to support the assignment. Citation(s) were accurate but may not have been formatted according to APA guidelines.	Cited a minimum of two peer-reviewed, scholarly, or similar references to support the assignment. Citations were accurate and formatted according to APA guidelines with few format errors.	Cited more than two peer-reviewed, scholarly, or similar references to support the assignment. Citations were accurate and formatted according to APA guidelines with no format errors.

APPENDIX F

BENCHMARK ASSIGNMENT STUDENT WORK SAMPLE

164

<div align="center">Problem</div>

I am the manager of Happy Cardiology services and we have been informed by human resources of some conflicts that have arisen within the organization. I've been asked to implement a program to address the concerns with the staff member of my team. There are concerns of coding, billing and credentialing staff. The staff feel that administration does not fully understand their job duties and does not provide a way for the staff to share their concerns about the work environment. There are billing claims that some pressure to up the codes and aggressively seek payment from patients. On the other hand, we face a problem of having low morale, staff resigning at a rapid pace which results in a turnover rate at 22%.

<div align="center">Problem and Solution</div>

It is my job as the manager of Happy Cardiology services to create and implement program to fix and build a plan which address the concerns of the staff's issues. Create a plan of better communication and training of staff to improve the administration duties. Give the remaining staff a platform to where they can share their concerns about the work environment. I would place a communication box in the breakroom for employees to write out their concerns and give them a chance to express their concerns in written form. Training staff is also a big portion of the solution to our problem. Staff will attend a mandatory employee and training credentialing program.

<div align="center">Why I Think these solutions will effective?</div>

The positive effects to letting employees express their concerns with the communication box is that it will allow people to express themselves freely, anomalously without penalty, this will allow upper management see and hear their concerns with

<div align="center">165</div>

structure, plans, visions, and management duties. The positive effects to training staff will be encouraging staff to go back to school to advance their education in the field and to learn about other positions within the company which will allow then to grow within them organization.

<u>Implementing the solution</u>

When implementing the solution to the problem, it's going to take strategic planning. This is going to require finding the best steps to finding out the concerns of the staff and having leaders, managers, directors to take those concerns and organize them the best way for staff to see the vision, see what will help them and benefit the company as well. We have an obligation to inform patients prior to the doctor's visit, and to address co-pays, and allow financial planning. We need systems back up to help send out billing and collections options. Accurate coding and billing is necessary to ensure correct costs and coverage. Billing and staff credentialing are needed to show evidence of correct documentation, diagnosis and payments to ensure no fraudulent practice.

<u>Leadership style that best fit the solution</u>

I believe the best form of leadership for our office is the democratic leadership. Democratic leadership is grounded on participation of leaders and staff. The role of leadership is Laissez-faire. This leader-ship style allows the staff to achieve control through less obscure means. It is believed that employees excel when left alone to respond to their responsibilities and obligations in their own way. This can help expand their performance capacity. In the end the leader chooses what is best for the organization insight of the team.

<u>Leadership roles in conflict management</u>

166

As mentioned above managing conflict is tough but it's important to remember to accommodate, stay away from avoiding certain issues, collaborate to strengthen commitment, compete important tasks that create wellbeing of the company and compete against people who take advantage of the weak. We need to compromise toward goals collaborate when things aren't successful. Lastly if problem solving doesn't results with the group leaders willingly will make the best decision for the group. (University of Phoenix, 2017 p 120)

Creating Effective work group for implementation of proposed solution

According to Willful Choice decision-making models, "Identify the problem, collect data, list all possible solutions, test possible solutions and select based on the decision made." (University of Phoenix, 2017 p.139) This requires a team effort, which includes what team members have gathered in the communication box. Employee meetings and one on one interviewing between staff and management are important. Management they will take the Reality-based decision-making model. According to Reality-based decision making models, "Researchers have observed that willful choice models of decision making underestimate the chaotic nature and complexity regarding actual decision-making situations; a large percentage of decision- making processes are followed without actually solving anything." (University of Phoenix, 2017 p 249.) I think this style of leadership will help Happy Cardiology Services.

In Conclusion

It is important to remember that in order for a successful organization Happy Cardiology needs collaboration and team work to be successful. Having to problem solve, find effective solutions and implementation using appropriate leadership style will help

167

reach resolutions and implementing the best plan to resolve conflict. An exceptional

leader or manager needs the above tools to run a successful program. Staff, leadership,

administrators need to collaborate effectively amongst themselves towards or attain

positive outcomes for Happy Cardiology to become a successful organization.

References:

Stefl, M.E., Robbins, Bradley, Spicer, & Meclenberg, J.H.E., M.A.B., &Boyatis, R.E. (2017). Leadership Competence: Professional Competenies, and Personal Skills and Responsiblites [University of Phoenix Custom Edition ebook]. Phoenix, AZ:. Retreived from University of Pheoenie=x, HCS/475 Leadership and Peformance Development website.

University of Phoenix. (2017). Leadership Competence II: Application of Skills, Tools, and Abilities. Retrieved from https://phoenix.vitalsurce.com/#/books/9781284046496/cfi/6/6!/4/2/2@0:0

APPENDIX G

BENCHMARK ASSIGNMENT – SUMMARY MEMO GUIDELINES

<div align="center">***Summary Memo Guidelines***</div>

Executive summary memos are sometimes the first and only thing executive management will read and use to decide if they support your plan.

Traditionally, summary memos are only one to two pages long; thus, challenging the writer to not only grab the reader's attention but to synthesize information and convince the reader as to the merits of the analysis and recommendations.

Structural Guidelines:

Although this is a shorter assignment, you will need to organize your memo as you would a traditional paper.

- Introduce your problem in the introductory paragraph.
    - Capture your audience's attention.
    - Summarize the problem, its significance, and discuss what your proposed solution entails.
- Use simple headings within the memo.
    - For example, the introductory paragraph may have a header titled "Problem".
- The conclusion needs to summarize key points and identify next steps.
- References in APA format are required when using outside sources.

Stylistic Guidelines:

- Keep paragraphs short and concise.
    - Include key points that are critical to your analysis, decision, and change efforts.
    - Review and ensure that the various sections flow together and are not disjointed (coherence).
- Maintain a professional tone throughout the memo.
    - Use language appropriate for your target audience.
        - For this assignment, your audience is the internal executive management team.
        - Keep in mind that your audience will change based on the organization you work for.
    - Minimize the use of jargon, slang, and acronyms.

APPENDIX H

NORMING INTERVENTION – FACILITATOR GUIDED QUESTIONS

Rubric orientation

Can you clearly distinguish between each performance descriptor on this line?

Is there a clear ranking order to the performance descriptors on this line?

What specific language in the rubric aligns to evidence in the student work sample?

How well does this assignment description align to the rubric?

Summary questions

Looking at all the scores together here, what level of consistency would you say there is?

Does hearing your colleagues' perspective change your opinion?

Does having an analytic rubric like this make your life easier or harder, in terms of setting expectations for learning?

Do students, in your mind, value the transparency of having a rubric?

Do students benefit from its use?

APPENDIX I

SAMPLE SCREENSHOTS FROM ASYNCHRONOUS NORMING SITE

173
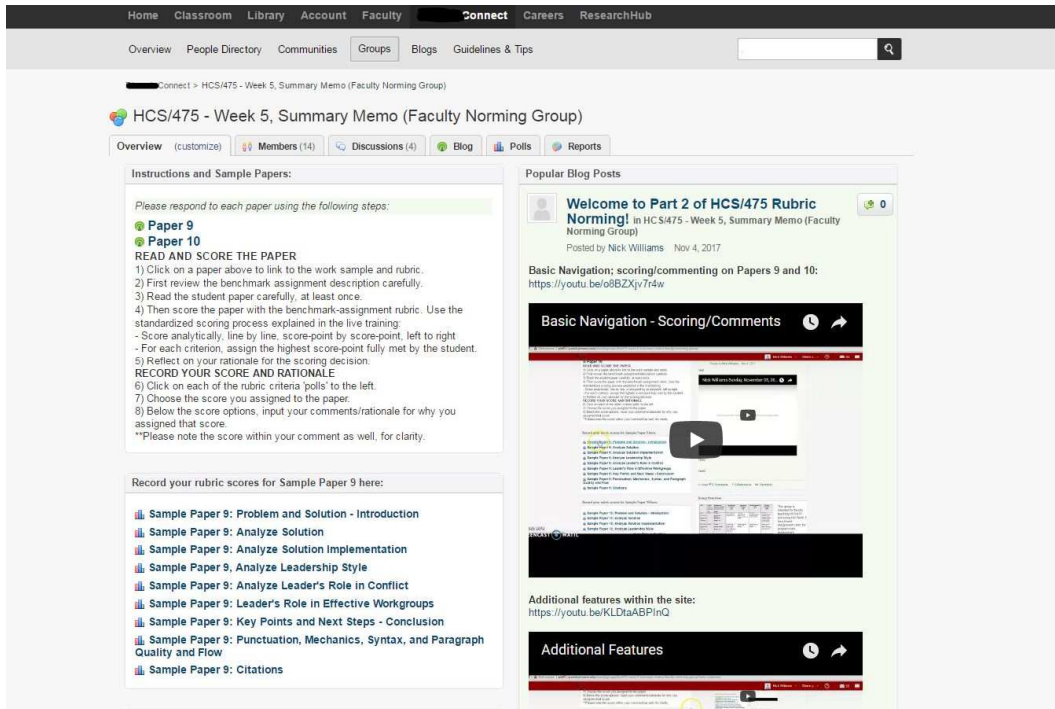
Figure A. Front page of the SWConnect asynchronous norming site. The two work samples and tasks are denoted on the top left-hand side. Bottom left are polls for participants to log scores. The right-hand side features how-to videos for navigation, basic tasks, and other features.

Figure B. Example of the scoring tool inside the site's "polling feature". Each rubric criterion has a poll. At the top is the actual rubric line with performance descriptors, beneath which participants log their vote or see anonymized voting results. Finally comments are provided by the participants to explain their scoring rationale.

175

Figure C. Example of a discussion post by facilitator and interaction among participants afterward.