

**A Comparative Case Study of Approaches and Institutional Factors that Affect
Assessment of Teaching Planning at a University**

by

Lindsay Onufer

Bachelor of Arts, Westminster College, 2004

Master of Arts, Point Park University, 2011

Submitted to the Graduate Faculty of the
School of Education in partial fulfillment
of the requirements for the degree of
Doctor of Education

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

SCHOOL OF EDUCATION

This dissertation was presented

by

Lindsay Onufer

It was defended on

November 10, 2021

and approved by

Lori Delale O'Connor, PhD, Assistant Professor, Center for Urban Education & Department of Educational Foundations, Organizations, and Policy

Michael G. Gunzenhauser, PhD, Associate Professor & Chair, Department of Educational Foundations, Organizations, and Policy

Charline Rowland, EdD, Associate Professor & Chair, Department of Education, Bethany College

Dissertation Advisor: Thomas Akiva, PhD, Associate Professor & Director of Schoolwide EdD, Department of Health and Human Development

Copyright © by Lindsay Onufer

2021

**A Comparative Case Study of Approaches and Institutional Factors that Affect
Assessment of Teaching Planning at a University**

Lindsay Onufer, EdD

University of Pittsburgh, 2021

For the past several decades, researchers have identified problems with the validity and reliability of student opinion of teaching survey (teaching survey) results, leading many researchers and faculty members to conclude that conducting comprehensive, meaningful assessment of teaching requires using multiple measures to collect and triangulate data from students, faculty peers, administrators, and others (AAUP, 1975; Arreola, 2007; Benton & Cashin, 2014; Berk, 2006; Vasey & Carroll, 2016). Despite the criticism of overreliance on teaching surveys, most institutions (Vasey & Carroll, 2016), including the University of Pittsburgh, continue using them as the primary means of assessing teaching effectiveness, and much is to be learned about how a university or academic unit can move to more comprehensive methods. In 2021, the University of Pittsburgh began an institution-wide process to create and implement plans to broaden and improve assessment of teaching. Using document analysis of assessment of teaching plan documents, I examined the approaches to and comprehensiveness of academic units' assessment of teaching plans. I conducted faculty focus groups to identify institutional factors that faculty perceived as having facilitated or impeded assessment of teaching planning. Results indicate that units that took team-based middle-out approaches, which required more faculty involvement than top-down, leader-led approaches, created more comprehensive plans. Focus group data analysis results also suggest that access to resources and aspects of unit culture affect this type of institutional change. Institutional drivers and barriers were also context-

specific at the unit-level. This study concludes with recommendations for how various stakeholders at the University of Pittsburgh and change agents in other higher education institutions can facilitate assessment of teaching planning and improvement moving forward.

Table of Contents

| | |
|--|-----------|
| Preface | xi |
| 1.0 Introduction | 1 |
| 1.1 Problem Area | 1 |
| 1.2 Organizational System | 3 |
| 1.3 Stakeholders | 6 |
| 1.3.1 Faculty | 7 |
| 1.3.2 Students | 8 |
| 1.3.3 Administrators/Unit Leaders | 10 |
| 1.3.4 Teaching Center Staff | 12 |
| 1.4 Purpose | 13 |
| 1.4.1 Feasibility and Implications | 15 |
| 2.0 Literature Review | 17 |
| 2.1 Best Practices in Assessment of Teaching | 17 |
| 2.2 Organizational Change in Higher Education | 21 |
| 2.3 Revising Assessment of Teaching as a Change Process in Higher Education | 25 |
| 3.0 Methods and Measures | 29 |
| 3.1 Document Analysis | 29 |
| 3.2 Focus Groups | 33 |
| 4.0 Results | 36 |
| 4.1 Analysis of Assessment of Teaching Plans | 36 |
| 4.1.1 Approach | 38 |

| | |
|--|----|
| 4.1.1.1 More Faculty Involvement in Middle-Out Plans | 38 |
| 4.1.1.2 More Context-Specific Definitions and Aligning Tools in Middle-Out Plans | 39 |
| 4.1.1.3 Use of Results for Improvement Beyond Individual Faculty Evaluations in Middle-Out Plans | 39 |
| 4.1.1.4 More Middle-Out Plans Addressed the Need for Reviewer Preparation | 40 |
| 4.1.2 Understanding of Expectations | 41 |
| 4.1.3 Flexibility | 41 |
| 4.2 Focus Groups..... | 42 |
| 4.2.1 Units' Processes for Revising Assessment of Teaching | 43 |
| 4.2.1.1 Middle-Out Group | 43 |
| 4.2.1.2 Top-Down Group | 44 |
| 4.2.2 Goals for Revising Assessment of Teaching | 44 |
| 4.2.3 Institutional Factors that Affected Assessment of Teaching Planning and Factors Necessary for Success | 49 |
| 4.2.4 How Revisions Will/Have Changed Assessment of Teaching | 55 |
| 5.0 Learning and Next Steps..... | 58 |
| 5.1 Discussion | 58 |
| 5.2 Next Steps and Implications..... | 63 |
| 5.2.1 At Pitt..... | 63 |
| 5.2.2 At Other Institutions | 69 |
| 5.3 Limitations and Suggestions for Future Research | 73 |

| | |
|---|-----------|
| 5.4 Conclusion | 75 |
| 6.0 Reflection | 77 |
| Appendix A Document Analysis Rubric..... | 81 |
| Appendix B Focus Group Interview Protocol and Questions..... | 83 |
| Appendix B.1 Protocol (adapted from Dos Santos Marques et al., 2020) | 83 |
| Appendix B.2 Questions..... | 84 |
| References..... | 86 |

List of Tables

| | |
|--|-----------|
| Table 1 Characteristics of Effective Assessment of Teaching with Aligning Performance | |
| Criteria | 31 |
| Table 2 Assessment of Teaching Plan Document Analysis Ratings by Performance Criteria | |
| | 37 |
| Table 3 Goals Theme..... | 46 |
| Table 4 Institutional Factors that Helped Assessment of Teaching Theme | 50 |
| Table 5 Institutional Factors that Hindered Assessment of Teaching Theme | 53 |
| Table 6 Institutional Factors that Will Be/Are Necessary for Assessment of Teaching | |
| Success Theme | 54 |
| Table 7 How Revisions Will/Have Changed Assessment of Teaching Theme | 57 |
| Table 8 Stakeholder Groups and Strategies for Support | 64 |

List of Figures

| | |
|--|-----------|
| Figure 1 ACE Project on Leadership and Institutional Transformation Typology of Change | 22 |
| Figure 2 Change Macro Framework..... | 23 |

Preface

I could not have completed this work without the support of my friends, family, colleagues, and mentors. To my parents, Lynne and Terry Onufer, thank you for nurturing my academic interests and for your unflagging encouragement. To my friends and colleagues, Katie Miller, Dr. John Radzilowicz, Dr. Lizette Muñoz Rojas, and Dr. Josie Rush, thank you for your patience, humor, and companionship.

I am also grateful for my advisor, Dr. Tom Akiva, who provided timely advice and expert guidance, including helping me navigate the daunting challenges that arose from completing dissertation research during the COVID-19 pandemic. Thank you to my committee members, Drs. Lori Delale O'Connor, Michael G. Gunzenhauser, and Charline Rowland, for your thoughtful feedback and support.

To Drs. Bethany Hicok, Diane Maldonado, Tahirah Walker, and Charline Rowland, the fierce, brilliant women who gave their time and wisdom so generously, thank you for the opportunities you created for me and the paths you forged.

Finally, to my partner, Shannon Fink, and the many animals we have rescued together, your love is everything to me.

1.0 Introduction

1.1 Problem Area

Debates about the usefulness and validity of various methods of assessment of teaching have proliferated in faculty meetings and publications for decades. The most widely used assessment of teaching measure is student opinion of teaching surveys (teaching surveys). According to a 2014 survey of 9,314 American higher education instructors conducted by the American Association of University Professors (AAUP), almost 100% of colleges and universities use some form of teaching survey to gather student feedback to improve teaching and inform evaluative decision-making like hiring and promotions (Vasey & Carroll, 2016). Despite their pervasiveness, in the nearly hundred years since universities adopted teaching surveys, few topics related to the evaluation of instruction have provoked as much faculty debate and criticism. From 2007 to 2017 alone, *The Chronicle of Higher Education* and *Inside Higher Education*, the two most popular higher education news publications in the United States, published over 50 articles about teaching surveys, 65% of which were negative essays and editorials (Linse, 2017).

Critics have contended that teaching surveys are overused, often as the only measure of teaching effectiveness, and produce flawed data that is, at best, only moderately correlated to student learning (American Sociological Association, 2019; Berk, 2006; Burdsal & Harrison, 2008; Esarey & Valdes, 2020; Gormally et al., 2014; Nasser & Fresko, 2002; Uttl et al., 2017; Vasey & Carroll, 2016). Studies have also shown that students' biases affect the teaching survey results of instructors who are members of minoritized groups, including women, people of color (particularly Black instructors), and international instructors (Fan et al., 2019; Mengel et al., 2019;

Mitchell & Martin, 2018; Peterson et al., 2019; Smith & Hawkins, 2011; Smith & Johnson-Bailey, 2012; Storage et al., 2016; Taylor, 2021). Overreliance on this single data source fails to produce robust evidence of teaching effectiveness (AAUP, 1975; Arreola, 2007; Benton & Cashin, 2014; Berk, 2006). Further, using a potentially invalid, biased measure to assess teaching can place minoritized faculty at an unfair disadvantage during formal evaluations and discourage faculty from using assessment data (Taylor, 2021), which may stymie continuous improvement of teaching.

Conducting teaching surveys does provide faculty with feedback from students, the group most directly affected by teaching. However, students lack the content and pedagogy expertise of faculty. Because of this, experts have recommended that teaching surveys should serve as one, ideally formative, assessment of teaching amongst several measures, including peer review, self-assessment, and review of teaching and student learning artifacts (AAUP, 1975; Arreola, 2007; Benton & Cashin, 2014; Berk, 2006; Burdsal & Harrison, 2008; Gormally et al., 2014; Linse, 2017; Miller & Follmer Greenhoot, 2018; Murphy et al., 2009; Nasser & Fresko, 2002). Faculty have also reported that they prefer to use teaching surveys in conjunction with other measures. In the 2014 AAUP instructor survey, 69% of the respondents stated that they saw the value in conducting teaching surveys, but 50% reported that they did not believe that teaching surveys alone were a good measure of teaching effectiveness (Vasey & Carroll, 2016). Collecting multiple types of evidence, including student feedback, enables faculty to compare and triangulate data from different sources to create a more complete overview of teaching effectiveness.

Despite the decades-long near-consensus amongst higher education organizations, researchers, and faculty that multiple sources of data are better than one, the overuse of teaching surveys persists in many institutions, including, until 2019, the University of Pittsburgh (Pitt).

Beginning in 2019, in response to faculty concerns about the use of teaching surveys (Barlow, 2017; Harrell, 2018), the provost charged the Advisory Council on Instructional Excellence (ACIE), an interdisciplinary committee of teaching award-winning faculty who work to advance instructional improvement and innovation, with composing recommendations to guide improvement of assessment of teaching. In 2021, the Office of the Provost directed academic units to use ACIE's recommendations to create plans to revise assessment of teaching.

1.2 Organizational System

Pitt is a large, public, research-intensive doctoral university with approximately 33,000 students and 5500 faculty (Office of Institutional Research, 2021). The main campus in Pittsburgh consists of 14 schools or academic units. There are also four regional campuses. According to Pitt's 2021-2025 strategic plan, the Plan for Pitt, the university's mission is to "leverage knowledge through teaching, research, and community service-for society's gain" (p. 4). Strategic goals focus on improving academics, research, the community, and opportunities for community members, including faculty, staff, and students (Plan for Pitt, 2021).

Although the mission and the Plan for Pitt indicate that the university values teaching effectiveness, Pitt's decentralized organizational structure and promotion and tenure policy permit flexibility regarding how academic units assess and reward teaching (Office of the Provost, 2020). This has led to significant variations in how units have evaluated teaching. For instance, in some academic units, instructors have been assessed using multiple measures for years; teaching survey results were one data point amongst several. In others, teaching survey results were the only method for assessing teaching. In 2019, ACIE created seven research-based recommendations for the revision of assessment of teaching with support from the University Center for Teaching and Learning (Golden & Kirsch, 2020, pp. 9-10):

1. Each academic unit would develop processes, criteria, and an action plan for evaluating teaching and submit the plan to the Office of the Provost by the end of the 2021 spring semester.
2. Plans should address the use of multiple sources of teaching effectiveness evidence, including student feedback, peer feedback, representative teaching materials, and self-assessment.
3. Plans should describe how faculty have been and would remain engaged in the process.
4. The university's Center for Teaching and Learning would support plan creation by providing resources and consultations.
5. Plans should describe strategies units used to inform faculty about best practices related to teaching surveys.
6. Units would develop guidance on the expectations and the use of data in evaluative processes like annual, promotion, and tenure reviews.
7. All deans and campus presidents would receive access to their unit's teaching survey results if they did not already have access.

The recommendations were discussed and endorsed by the Council of the Deans, then shared with the Faculty Senate Educational Policies Committee, and accepted by the Faculty Senate in November of 2020. The Office of the Provost released a memo on changes to assessment of teaching on February 16, 2020. In March 2020, Pitt shifted to remote instruction in response to the COVID-19 pandemic, which delayed further progress for a year. In early spring 2021, the Office of the Provost asked academic units to submit plans by April 30, 2021, the end of the spring semester.

My place of practice, the University Center for Teaching and Learning (Teaching Center), is an approximately 70-staff unit led by an associate provost/executive director. The Teaching Center acts as a centralized teaching support resource for all academic units in the university and houses sub-units with varied goals. Teaching Center sub-units help faculty improve teaching, integrate educational technology into their teaching, design online courses, borrow classroom equipment, and support other types of projects related to teaching and learning. Pitt's Teaching Center is one of the largest in the United States and has a centralized, top-down organizational structure. The associate provost creates policies and operational guidelines, which are communicated through directors to managers and staff. The mission of the Teaching Center is to "inspire excellence and innovation in teaching, learning and scholarly activities at the University of Pittsburgh" (University Center for Teaching and Learning, 2021). Departmental strategic goals mirror the goals in the Plan for Pitt.

The Teaching Center contains an initiative and an office that support assessment of teaching: The Assessment of Teaching Initiative and Office of Measurement and Evaluation (OMET). The Assessment of Teaching Initiative supports individual faculty and academic units in planning, conducting, interpreting, and using assessment for teaching improvement. OMET manages, deploys, and reports on teaching surveys. My position, program manager of the Assessment of Teaching Initiative, was created in 2019 in response to the provost's and faculty's growing interest in assessing teaching effectiveness. I have expanded assessment of teaching resources and support services, assisted academic units with creating assessment of teaching plans, provided units with feedback on submitted plans, and am currently supporting plan implementation. Because many academic units relied heavily on teaching surveys to evaluate faculty teaching, a significant portion of my work has involved raising awareness about and

helping in the creation of other measures of teaching, particularly peer observation and review tools.

Assessment of teaching aligns with the department's mission and strategic goals. Creating my position and the initiative indicates that the Teaching Center leaders support assessment of teaching efforts at Pitt. However, like other universities, Pitt's budgetary constraints were exacerbated by the COVID-19 pandemic, leaving the Teaching Center unable to allocate resources for growing the Assessment of Teaching Initiative further. Lack of funding also affects staffing. Most staff who left or took early retirement during the pandemic have not been replaced, leaving remaining staff to absorb their duties. The Teaching Center's service model, which is primarily reactive rather than proactive, also limits the Assessment of Teaching Initiative's reach. Although ACIE and the Office of the Provost have encouraged academic units to consult with the Assessment of Teaching Initiative, I can only work with units and faculty who request my assistance.

1.3 Stakeholders

Research literature, my professional experience, and interviews conducted with five instructors and one student confirmed that four groups of stakeholders are invested in this issue: faculty, students, administrators/unit leaders, and Teaching Center staff.

1.3.1 Faculty

Empathy interviews are a design thinking, improvement science methodology which allow researchers to collect rich feedback from stakeholders within a system to explore a problem (Valdez et al., 2020). To better understand Pitt faculty perspectives on this issue, I conducted 1.5-hour, semi-structured empathy interviews with five University of Pittsburgh faculty in 2019. I recruited interview participants from groups of faculty who had worked with the Teaching Center on teaching initiatives to ensure the collection of rich, detailed data from participants knowledgeable about assessment and improvement of teaching. Participants each taught in different disciplines and represented different rank and appointment levels ranging from part-time faculty to full-time, tenured faculty, but shared a commitment to improving teaching and how it is assessed. All the instructors interviewed identified as members of minoritized identity groups.

Common themes emerged from interviews. Faculty expressed that they:

- valued student feedback and using it to improve their teaching;
- valued assessing teaching, but believed that the ways their schools and departments evaluated teaching were insufficient and flawed;
- believed that teaching surveys could be better designed and should not (but often were) used as the only means of evaluating teaching;
- had all experienced receiving biased teaching survey results.

The degree to which academic units weighted teaching survey results and instructors' perception of the quality of teaching survey data varied. Some felt that survey data was generally valid and reliable. Others questioned the quality of teaching survey data due to students' lack of pedagogy expertise and because results reflected students' biases. One faculty member stated that

they believed biased survey results create systemic issues, including perpetuating institutional inequity and inhibiting teaching innovation:

I feel like we're constantly worried about [teaching surveys]. I feel it really impacts any type of experimentation you'd like to do with teaching. And that's particularly damning for female faculty of color because they're the ones who receive more harsh criticisms and it sort of perpetuates this idea of an elite institution and unfortunately, that they don't belong there because they're not as good. So, I think it perpetuates gender bias and race bias and it's used to profoundly discredit the work of professors. And to constantly have that in the back of your head is really difficult when you try to organize a new class. (interview, December 10, 2019)

While these initial interviews do not constitute a representative sample of faculty, themes from interviews and research on assessment of teaching were similar. The instructors did not indicate that they were opposed to conducting teaching surveys. Rather, most stated that they would prefer more comprehensive, equitable assessment of teaching conducted by trained reviewers using multiple methods and sources of evidence.

1.3.2 Students

Pitt's approximately 33,000 students primarily participate in assessment of teaching by completing teaching surveys. For many students, completing teaching surveys is the only time during a semester when they give formal feedback on instructional effectiveness, but the timing and misuse of teaching surveys can be problematic. Waiting until the end of the semester to collect student feedback increases the likelihood that end-of-semester stress influences students' responses. By the time faculty receive teaching survey results, it is too late for them to make

changes to improve the learning experiences of the students who offered the feedback. As instructors indicated in empathy interviews, students also lack the pedagogy or content expertise to evaluate teaching, although their feedback is often used in faculty evaluations (Benton & Cashin, 2014; Gormally et al., 2014; Linse, 2017).

In a 2007 survey administered to approximately 600 students across 20 institutions (Campbell & Bozeman), two thirds of students agreed that conducting teaching surveys was useful and almost all reported answering teaching survey questions honestly. However, many factors may influence how students rate instructors. An undergraduate student I interviewed reported that they base their survey responses on the quality of the instruction they received, factors related to their experience in the course but not related to instruction (like class time or location, for example), and their personal feelings about the instructor or course (personal interview, December 4, 2019).

The degree to which students understand the purpose and use of teaching survey data is also unclear. Previous studies have indicated that students may be unaware of how teaching survey data contributes to formal faculty evaluation processes (Campbell & Bozeman, 2007). At Pitt, outside of OMET emails containing teaching survey reminders and links, students do not receive any information about assessment of teaching policies or how teaching survey data are used unless instructors offer that information. Some students might understand teaching surveys as a customer satisfaction survey. Others might approach it as an evaluative procedure. Depending on the instructor's academic unit, either could be correct. Lack of communication about how teaching surveys are used only further complicates how students might engage with surveys. The student I interviewed stated that, in their experience, many instructors do not even mention teaching surveys or, if they do, they simply remind students to complete them. The student was not sure of the purpose of teaching surveys, not confident that their instructors read results, and was unaware of

whether teaching was evaluated in any other ways at Pitt (personal interview, December 4, 2019), which is largely consistent with previous studies of student perceptions of teaching surveys (Campbell & Bozeman, 2007). Given students' varied perceptions and uses of teaching surveys, the degree to which they value them likely differs.

Graduate students are also members of this stakeholder group and complete teaching surveys for their instructors. In addition, graduate teaching assistants, depending on their academic unit and role, may be evaluated themselves using teaching surveys. Because the amount and type of pedagogical training and mentoring they receive varies, some graduate students may not be aware of how teaching surveys are commonly used, how to interpret and apply teaching survey data to improve teaching, or of the potential limitations of survey results. To my knowledge, graduate students have no authority to propose changes to how teaching surveys are designed besides, in some case, having the ability to add questions to their teaching surveys. Despite potentially having little training on or control over teaching surveys as measures of teaching effectiveness, survey results could affect their job prospects as they enter the academic job market.

1.3.3 Administrators/Unit Leaders

Pitt's provost has consistently championed developing more comprehensive and effective assessment of teaching since 2019. By requiring academic units to submit assessment of teaching plans, she helped catalyze discussions about how to improve assessment of teaching throughout the institution. ACIE developed broad recommendations to guide assessment of teaching plan creation so academic units would have the flexibility and autonomy to compose plans best suited to the units' needs. Academic unit leaders have had different levels of involvement in revising assessment of teaching and creating plans. All the unit leaders with whom I have worked have

expressed a commitment to improving teaching in their unit; however, before developing assessment of teaching plans, most units overrelied on teaching surveys. As in many other higher education institutions (Fairweather, 2005; McMurtrie, 2019), formal faculty evaluation processes at Pitt also prioritize and reward research more than teaching, indicating a misalignment of espoused values and practices.

Even assuming that teaching survey data alone was a valid, unbiased measure of teaching effectiveness, research indicates that administrators and leaders are prone to misinterpreting results. A review of research on teaching data use practices (Linse, 2017) determined that administrators and faculty who examine teaching survey data may:

- fail to consider response rate and, thus, fail to recognize that low response rates make for less reliable results;
- fail to recognize when a drop in ratings may be due to an instructor implementing an innovative new teaching method;
- consider scores for a single semester or a single course in isolation rather than looking at trends across courses and over time;
- study mean quantitative ratings scores without looking at median scores, which increases the likelihood that a few extreme negative ratings could skew results;
- compare scores between instructors, which can be especially problematic in academic units with many skilled teachers. Comparing one highly effective instructor to another may create the perception that one is less skilled when the difference in scores is actually negligible.

Several of the instructors I interviewed reported experiencing these types of misinterpretations.

Two instructors suggested that unit leaders purposefully misinterpret teaching survey data to

prevent minoritized instructors from advancing in their careers. Whether accidental or purposeful, the prevalence of misinterpretations of teaching survey data suggests that data generated by other assessment methods like observations could also be interpreted incorrectly.

In working with units creating assessment of teaching plans, some units have indicated that they are aware of the potential of bias or misinterpretation to affect other measures of teaching effectiveness. Some units have planned to create reviewer trainings to mitigate these issues, but many have not. In academic units where misinterpretation or misuse of assessment of teaching data is more prevalent, faculty may trust unit leaders and assessment of teaching data less. Unit leaders' attitudes and use of data may also shape unit culture and the extent to which faculty value teaching and allocate time to teaching improvement.

1.3.4 Teaching Center Staff

The final group of stakeholders are the Teaching Center staff who support teaching improvement efforts. Except for the OMET manager and I, other Teaching Center staff are less directly connected to assessment of teaching than other stakeholder groups. Teaching Center staff have little influence over institutional assessment of teaching policies. Our knowledge of assessment of teaching activities or data is limited to what academic units and faculty share with us. Staff in the Teaching Support sub-unit facilitate formative assessment of teaching efforts (like teaching consultant observations) and faculty development to help instructors improve teaching. Because our jobs focus nearly entirely on supporting teaching, Teaching Center staff have the time and resources to prioritize and attend to teaching effectiveness that faculty and unit leaders may lack due to competing priorities.

Ideally, administrators, unit leaders, and faculty would view Teaching Center staff as experts they can rely on for teaching support and guidance. In reality, only a small percentage work with Teaching Center staff on assessment of teaching beyond using the teaching surveys that OMET deploys. In my experience, the leaders and faculty who proactively seek to collaborate with Teaching Center staff need our services the least. When academic units or faculty use our services, they generally find them helpful. For example, instructors who work with the Teaching Center for a year or longer tend to see increases in teaching survey scores. However, Teaching Center staff have limited control over who we work with and in what capacity. Consistent support from the Office of the Provost and increased faculty use of the Teaching Center during the COVID-19 pandemic may increase the trust in and demand for Teaching Center support moving forward.

1.4 Purpose

Until 2021, the assessment of teaching practices at Pitt were consistent with national trends. Many academic units only used teaching surveys to assess teaching, which was counter to assessment of teaching experts' recommendations (American Sociological Association, 2019, AAUP, 1975; Arreola, 2007; Benton & Cashin, 2014; Berk, 2006; Burdsal & Harrison, 2008; Gormally et al., 2014; Linse, 2017; Miller & Follmer Greenhoot, 2018; Murphy et al., 2009; Nasser & Fresko, 2002) and Pitt faculty's preferences (Barlow, 2017; Harrell, 2018). Research indicates that faculty members' concerns were warranted. Studies have found that teaching survey data is affected by student biases (Fan et al., 2019; Mengel et al., 2019; Mitchell & Martin, 2018; Peterson et al., 2019; Smith & Hawkins, 2011; Smith & Johnson-Bailey, 2014; Storage et al., 2016) and fail to predict effective teaching (Esarey & Valdes, 2020) or correlate to student learning (Uttl et al.,

2017). Using teaching surveys as the only measure of teaching effectiveness in hiring and promotion decisions is neither equitable nor comprehensive. As faculty participants reported in the empathy interviews I conducted, distrust in teaching survey results and leaders' abilities to interpret results properly discouraged some instructors from using them to improve their teaching or even reading them. Overuse of teaching surveys may have undermined the goal of assessment: improvement.

Research on organizational change in higher education suggests that some widely acknowledged problems like teaching survey overuse may be particularly pernicious due to the unique characteristics of higher education institutions (Boyce, 2003; Clark, 1983, Eckel et al., 1999; Weick, 1976). Universities consist of interconnected but somewhat autonomous units with their own cultures and values (Weick, 1976). The university leaders' and units' goals may or may not align, making enacting collective change difficult (Boyce, 2003; Clark, 1983, Eckel et al., 1999; Weick, 1976). Units in Pitt followed a standard set of recommendations to create an institution-wide change but planned and managed that change differently. Some units took a top-down approach (Cummings et al., 2005), defined as a centralized change management strategy led by leaders. Other units adopted a middle-out approach, meaning that the change had some support from leaders and faculty but required cultivating buy-in and collaborative decision-making (Cummings, et al., 2005; Reinholz et al. 2015; Weaver et al., 2020). In their study of middle-out approaches to revising teaching and learning, Cummings et al. (2005) proposed that middle-out approaches are led by "middle managers" (p. 11) who have more autonomy and access to resources than individual staff. Reinholz et al. (2015) established a model of systemic higher education change in which a middle-out approach involved a group of champions working to build broader departmental buy-in for a change initiative. In the assessment of teaching revision process at Pitt,

senior academic administrators like deans often delegated the assessment of teaching plan development to associate deans, department chairs, and faculty committee leaders and members. These groups of mid-level unit leaders and faculty worked together to build support for and advance assessment of teaching revision.

Although the structure of higher education institutions explains, in part, why implementing change can be challenging, it is unclear what specific factors influenced this particular change process and how. Given the dearth of research on institution-wide assessment of teaching revision, particularly in highly decentralized institutions, examining this change process at Pitt can help address several questions:

- When it comes to developing a comprehensive assessment of teaching plan, is one approach (top-down vs. middle-out) more successful than the other?
- What do faculty perceive to be the institutional factors (structures, processes, policies, aspects of institutional culture, and resources) that drive and impede assessment of teaching revision?

Answering these questions will help better define the elements necessary for and potential roadblocks to successful assessment of teaching revision.

1.4.1 Feasibility and Implications

Academic units at Pitt are in the early stages of revising assessment of teaching; most are developing tools and processes to pilot in the 2021/2022 academic year. The results of this dissertation of practice provide academic units with information that they can use to shape assessment of teaching work moving forward. As the program manager of the Assessment of

Teaching Effectiveness Initiative, I am well-positioned to propose recommendations based on the results.

This study also contributes to research on organizational change in higher education and, more specifically, the burgeoning body of research on instituting more comprehensive assessment of teaching. Most research on assessment of teaching has focused on the validity, reliability, and efficacy of methods of assessment like teaching surveys. There are few studies on universities' approaches to improving assessment of teaching. Much of the existing research comes from one project, the Transforming Higher Education – Multidimensional Evaluation of Teaching or TEval. TEval consists of a networked improvement community of three institutions that used a common, institutional-level framework, adapted slightly by each university, for defining teaching effectiveness and managing revising assessment of teaching (Andrews et al., 2020; Finkelstein et al., 2020; TEval, n.d.; Weaver et al., 2020). Even fewer studies address faculty perceptions of institutional factors that affect revising assessment of teaching. Examining an institution-wide, decentralized revision of assessment of teaching will address these gaps in the research literature and help universities determine how to best implement measures of teaching beyond teaching surveys. More widespread adoption of more robust measures of teaching could help establish more equitable and accurate evaluation of faculty teaching and incentivize teaching improvement.

2.0 Literature Review

Despite the ubiquity of teaching survey overuse in higher education, research on assessment of teaching offers a blueprint for assessing teaching more effectively and comprehensively. Envisioning what better assessment of teaching might look like is less challenging than determining how to plan and implement the changes needed to transition to more effective teaching assessment and improvement. In *How Colleges Change*, Kezar (2018) argued that there are rarely truly unprecedented changes in higher education. Nevertheless, leaders and change agents are prone to focusing on responding to problems with simplistic interventions and overlooking the complexity of change processes (Kezar, 2018). To attend to both the product and process of revising assessment of teaching, this literature review will address the following questions:

- What constitutes effective, comprehensive assessment of teaching?
- How does change occur in higher education organizational systems?
- How have other universities planned and implemented institution-wide assessment of teaching revision?

2.1 Best Practices in Assessment of Teaching

As early as 1975, the AAUP has argued for a more fair and comprehensive way to evaluate teaching, contending that,

Colleges and universities properly aspire to excellence in teaching. Institutional

aspirations, however, have not often led to practices that clearly identify and reward teaching excellence, and the quality of teaching is not in fact the determining consideration in many decisions on retention, promotion, salary, and tenure. (n.p.)

The AAUP (1975) outlined several tenets for improving assessment of teaching, proposing that higher education institutions should:

- Clarify policies on evaluation of teaching and rewards for effective teaching.
- Develop systematic ways of evaluating teaching effectiveness that do not rely on the subjective judgment of administrators, then clearly communicate evaluation criteria and processes to faculty.
- Conduct multiple measures of teaching effectiveness, which faculty should help select or develop. Methods of assessment should align with the specific context in which the faculty member is teaching, meaning that evaluation should take the discipline, course type, and method of delivery into consideration. It would also involve collecting multiple types of evidence, which might include student feedback, peer review, review of teaching artifacts, lists of courses taught, numbers of students, and teaching-related scholarship, service, innovation.
- Give faculty a primary role in evaluating one another.
- Ensure that faculty being evaluated would receive formative feedback and teaching development that supported their improvement.

The AAUP (1975) stated that following these tenets would improve the quality of teaching and contribute to continuous improvement.

In the nearly 50 years since the AAUP endorsed its “Statement on Teaching Evaluation,” other seminal guides on creating faculty evaluation systems have reiterated these recommendations

and expanded on implementing them. In *Thirteen Strategies to Measure College Teaching*, Berk (2006) argued for the adoption of the American Educational Research Association's, the American Psychological Association's, and the National Council on Measurement in Education's 1999 Joint Standards for Educational and Psychological Testing to establish standardized measurement processes and to increase the rigor of assessment by applying criterion-referenced measures to teaching. In *Developing a Comprehensive Faculty Evaluation System*, Arreola (2007) echoed these sentiments, noting that accrediting bodies increasingly requiring comprehensive measures of competency caused a paradigm shift toward increased rigor and accountability in higher education. Institutions could no longer rely on the evaluation based on unspecified criteria; they must generate robust evidence of teaching effectiveness using comprehensive measurement of faculty performance, including teaching (Arreola, 2007).

Enacting this type of systematic evaluation of teaching would necessitate developing context-specific definitions and characteristics of teaching effectiveness, conducting multiple measures for different purposes, then triangulating data to arrive at a complete picture of teaching quality (Arreola, 2007; Berk, 2006). Measures should consist of various types of ratings (from the faculty member, students, peers, alumni, administrators, and/or employers hiring students), teaching observations, teaching awards and scholarship, teaching portfolios, and/or student learning outcomes measures (Berk, 2006). Arreola (2007) ceded that, although drawing data from multiple sources, including fellow faculty, would create more complete evidence of teaching effectiveness, ensuring complete objectivity of any measure of teaching would not be possible. Instead, he argued that assessment of teaching should aim for "controlled subjectivity" (p. xix), meaning that a consistent set of performance criteria should be applied to teaching effectiveness data in consistent ways. The act of conducting multiple measures in and of itself is not sufficient

to ensure validity, reliability, and bias-free data, but training reviewers to analyze data correctly and using several sources of evidence would reduce the likelihood that biased data from a single measure would significantly hurt a faculty member during an evaluation (Arreola, 2007; Berk, 2006).

Although faculty tend to support improving teaching assessment (Vasey & Carroll 2016), the process of developing and implementing multiple, rigorous measures of teaching effectiveness is still challenging. In a 2008 AAUP presentation on improving faculty teaching (Miller & Follmer Greenhoot), the presenters argued that improving assessment of teaching requires a cultural shift in higher education institutions. Creating successful assessment systems necessitates widespread, vocal support from institutional leaders and consistent support and engagement from faculty (Arreola, 2007; Berk, 2006; Miller & Follmer Greenhoot, 2008). Arreola (2007) noted that faculty might resist new assessments of teaching despite widespread faculty criticism of teaching survey overuse if they are anxious about having not received adequate teaching training, resent the implication that their teaching skills could be questioned, fear accountability, or distrust the competence and objectivity of peers and leaders conducting assessments. Faculty could also be ambivalent to improving teaching, particularly when research accomplishments are often rewarded more than teaching excellence (Arreola, 2007). In addition to faculty support, managing this type of change requires commitment from leaders to provide faculty with the resources, like access to professional development, and incentives that drive improvement (Arreola, 2007; Miller & Follmer Greenhoot, 2008). Although assessment of teaching researchers have pointed out the problems with overusing teaching surveys, they also acknowledge that improving assessment of teaching constitutes a major and potentially difficult organizational change (Arreola, 2007; Berk, 2006; Miller & Follmer Greenhoot, 2008).

2.2 Organizational Change in Higher Education

Researchers have long recognized the challenge of understanding and managing change in higher education, particularly profound change. Compared to other systems, universities are often organized in a less centralized manner and are likelier to use a shared governance system to plan changes and make decisions (Storberg-Walker & Torraco, 2004). The ways in which units within the system function affect system-wide change (Clark, 1983; Weick, 1986). Weick (1976) described educational organizations as “loosely coupled systems” (pp. 6-8), or units that are connected within the same system by some shared goals or variables but maintain individual identities. Units develop values, culture, goals, and authority that may align or conflict with other units and the university’s administration (Boyce, 2003; Clark, 1983; Eckel et al., 1999; Kezar, 2018; Storberg-Walker & Torraco, 2004).

Managing change in loosely coupled systems has unique benefits and challenges. Individuals can act with more autonomy, and the system can create and sustain change at the unit-level more easily (Weick, 1976). Problems can also be isolated at the unit-level and are less likely to spread institution-wide (Weick, 1976). However, higher education institutions may have less focused, more ambiguous goals (Clark, 1983). They cannot adapt to or make institutional-level changes as quickly and efficiently as tightly coupled systems (Storberg-Walker & Torraco, 2004; Weick, 1976). Decision-making is slower and more diffuse (Clark, 1983; Weick, 1976). When planning institution-wide change, educational systems must balance the potentially competing needs and interests of units and stakeholder groups like students, faculty, staff, and the administration (Boyce, 2003; Clark, 1983; Eckel et al., 1999; Kezar, 2018; Storberg-Walker & Torraco, 2004).

Because of the structure of universities, research indicates that change agents must examine the context in which the change will occur and determine the type or level of change (Eckel et al., 1999; Kezar, 2018). One of the most influential guidebooks on managing change in higher education emerged from the American Council on Education (ACE) Project on Leadership and Institutional Transformation, which documented change initiatives at 26 institutions over the course of six years (Eckel et al., 1999). ACE researchers who supported these change initiatives identified a three-phase process to enact change: creating the context, developing change strategies, then marking progress and providing evidence of change. In the first phase, change agents focused on creating a context for change by determining the type of change, who should be involved, the decision-making structure of the institution, and institutional culture and attitudes toward change. They differentiated types of change by depth and pervasiveness (see Figure 1: ACE Project on Leadership and Institutional Transformation Typology of Change).

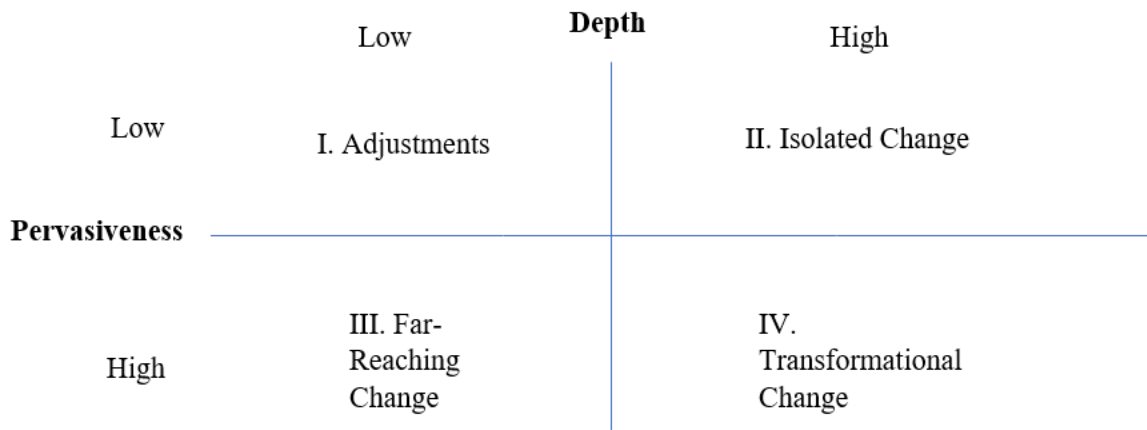


Figure 1 ACE Project on Leadership and Institutional Transformation Typology of Change

(Eckel et al., 1999, p. 21)

They determined that the deepest and most pervasive changes, which they described as transformational, required more time, planning, and sometimes involved overcoming significant obstacles.

Drawing on theory and research from dozens of studies of organizational change in higher education, Kezar (2018) proposed a similar but more detailed and comprehensive framework for developing approaches to change. Using this framework (see Figure 2: Change Macro Framework), change agents should determine the type of and context for change, their own authority to enact change, and ultimately develop change strategies drawing from relevant theories of organizational change.

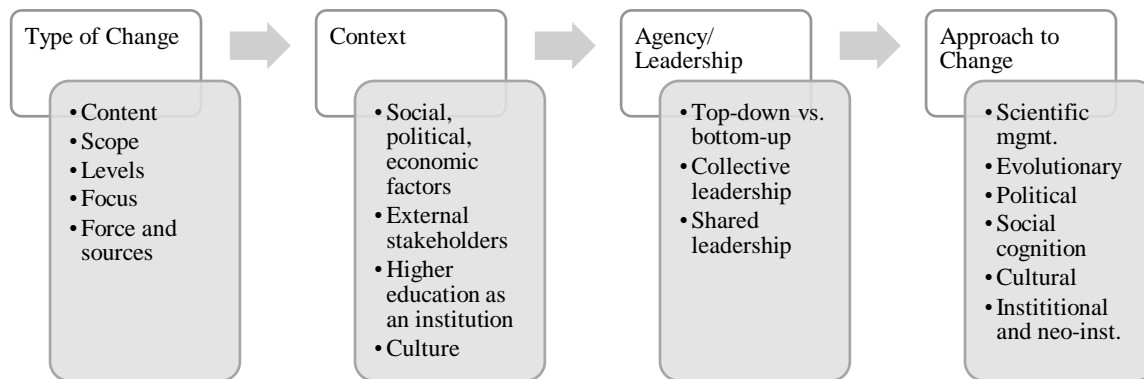


Figure 2 Change Macro Framework

(Adapted from Kezar, 2018, p. 134)

Kezar (2018) distinguished between first and second-order change. First-order changes resemble what the ACE Project on Leadership and Institutional Transformation described as adjustments. They are incremental structural and procedural changes that fit within an existing university system. Second-order changes are transformational and necessitate organizations critically examining and changing their cultures, assumptions, or values.

Kezar (2018) also highlighted the role of organizational learning in enacting change. Organizational learning derives from social cognitive theories of change which propose that

universities are social systems in which individuals' thoughts and learning shape change. Argyris (1999), the author of many foundational texts on the topic, described organizational learning as a process of identifying and correcting organizational problems that prevent the organization from learning and progressing towards achieving its goals. Argyris (1999) defined two types of organizational learning: single loop and double loop. When individuals engage in single loop learning, they examine organization problems and adjust their actions to address the problems. To engage in double loop learning, individuals must address underlying norms, assumptions, policies, and practices that cause the problem. Second-order change requires double loop learning (Boyce, 2003; Kezar, 2018).

Successfully enacting change, especially transformational or second-order changes, that challenge the cultural norms of the institution, requires change agents to anticipate resistance (Eckel et al., 1999; Kezar, 2018). Change strategies that address challenges and sources of resistance include (Eckel et al., 1999; Kezar, 2018):

- Establishing teams to work collaboratively on change tasks or strategies. Teams can make decisions and respond to changes faster. Members can contribute diverse knowledge and skills and represent the interests of various stakeholder groups.
- Engaging the campus community by cultivating buy-in and consensus, strategically crafting and communicating messaging, and soliciting ideas and feedback on change.
- Developing processes for and allocating resources like funds, time, and attention to instigating and sustaining the change.

Beyond building the infrastructure and capacity for change, higher education institutions must plan to unite units in a shared purpose. The primary challenges of changing the way universities assess teaching are not changing processes and tools. Rather, it is the transformational,

second-order change that leaders and faculty must make to examine their assumptions, values, norms, and behaviors related to teaching (Arreola, 2007; Miller & Follmer Greenhoot, 2008).

2.3 Revising Assessment of Teaching as a Change Process in Higher Education

In 2017, the AAU released *Aligning Practices to Policies: Changing the Culture to Recognize and Reward Teaching at Research Universities* (Dennin et al.), a report calling on universities to develop more effective ways to evaluate and incentivize teaching, particularly in STEM disciplines where teaching improvement lagged. The authors examined 51 institutions' promotion and tenure policies. They determined that, while most indicated that the university valued teaching, fewer contained guidance on measuring teaching or defined teaching effectiveness. They surveyed over 1000 instructors and asked respondents how their institutions value and reward teaching excellence. Most respondents agreed that their department leaders valued teaching improvement, but fewer agreed that teaching effectiveness played a meaningful role in formal evaluation processes. The report concluded with three case studies from UC Irvine, the University of Colorado, and the University of Kansas, institutions that constructed models for better teaching evaluation. UC Irvine revised weighting faculty research, teaching, and service to establish minimum standards for each of these three aspects of faculty work and evaluate each separately. The Universities of Colorado and Kansas had begun developing institution-wide frameworks for assessing teaching.

The Universities of Kansas and Colorado eventually joined with the University of Massachusetts to collaborate with the AAU in a networked improvement community (NIC) seeking to improve teaching assessment. The project, Transforming Higher Education –

Multidimensional Evaluation of Teaching (or TEval), has engaged in rapid, iterative testing to generate models for and research on change processes for transforming assessment of teaching. All three institutions took a similar, centrally-supported approach by first developing a standard framework and rubrics, which were adapted slightly by each institution (Andrews et al., 2020; Finkelstein et al., 2020, TEval, n.d.; Weaver et al., 2020). The Teaching Quality Framework (TQF) consists of seven dimensions of teaching (Andrews et al., 2020; Finkelstein et al., n.d.; Finkelstein et al., 2020):

- goals, content, and alignment;
- teaching practices;
- achievement of learning outcomes;
- class climate;
- reflection and iterative growth;
- mentoring and advising.

The framework also indicates that universities should review multiple sources of evidence to examine these seven dimensions of teaching.

Andrews et al. (2020) described the three-phase process for implementing the TQF. During phase 1, change agents assessed the readiness of various stakeholder groups for change and cultivated interest in the change effort by engaging in discussions with leaders, faculty in departments, and other stakeholders; determining whether they could build on existing change efforts; addressing sources of external pressure like accreditation standards and internal concerns like faculty anxiety over opaque evaluation processes. In the second phase, change agents formed teams made up of various stakeholders, including department leaders, faculty on relevant committees, and faculty champions passionate about improving teaching. The final phase

consisted of an iterative cycle in which a central support office or team helped teams identify forms of evidence, develop assessment processes, implement new assessment measures, and revise them based on data and feedback from departments (Andrews et al., 2020). At the Universities of Colorado, Boulder; Kansas; and Massachusetts, Amherst, 19, 12, and 9 departments respectively have adopted and adapted the Teaching Quality Framework and model to assess teaching (Finkelstein et al., 2020).

The TEval NIC (Finkelstein et al., 2020; Weaver et al., 2020) identified barriers in this change process consistent with challenges described in the literature on organizational change in higher education. Faculty champions within departments struggled to cultivate and maintain buy-in from colleagues and leaders. This reflects the challenge of fostering widespread support for a transformational change that requires a cultural shift (Eckel et al., 1999; Kezar, 2018). Faculty also had difficulty learning how to apply the rubrics derived from the Teaching Quality Framework (Finkelstein et al., 2020; Weaver et al., 2020), which indicates the need for infrastructure and resources to support implementing new processes (Eckel et al., 1999; Kezar, 2018). The TEval NIC (Finkelstein et al., 2020; Weaver et al., 2020) emphasized the importance of examining institutional culture, values, and history or establishing a context for the change (Eckel et al., 1999; Kezar, 2018). Coalition-building and seeking support from leaders and champions in the department also advanced assessment of teaching improvement (Finkelstein et al., 2020; Weaver et al., 2020).

Taken together, literature on organizational change and assessment of teaching suggests that improving the way institutions evaluate teaching requires multi-level change. First-order changes consist of adjustments to structures, processes, and tools used to measure teaching. Second-order changes consist of modifying how institutions that have traditionally valued research

more than teaching (Fairweather, 2005; McMurtrie, 2019) consider, incentivize, and reward teaching. The structure of universities indicates broad change must be managed and sustained at the institutional and unit levels. Team-based, middle-out approaches may be more effective than top-down approaches because faculty within the units can build support and develop strategies to advance change that take unit culture into consideration (Andrews et al., 2020; Eckel et al., 1999; Finkelstein et al., 2020; Kezar, 2018; Weaver et al., 2020). The research literature also points to potential institutional factors that could drive or thwart making this change: Unit culture and the degree to which revising teaching assessment aligns or misaligns with unit values and norms; unit leader and faculty support or resistance; and access to resources, including funds, time, and attention.

3.0 Methods and Measures

The aim of this dissertation of practice is to determine whether a top-down or middle-out approach is more effective when it comes to developing assessment of teaching plans and to identify and describe the institutional factors that faculty perceive to affect assessment of teaching revision. To address this aim, I conducted a comparative case study to examine units' approaches, comprehensiveness of assessment of teaching plans, and faculty perceptions of institutional factors. Case study methodology is used for collecting rich data to examine phenomena and how and why they happened (Yin, 2014). Comparative case studies are used to identify similarities, differences, and patterns across cases to answer research questions (Goodrick, 2014; Zartman, 2005). Using a comparative case study methodology allowed me to identify patterns and differences in assessment of teaching plan comprehensiveness and institutional factors and how they affected plan creation. The comparative case study consisted of two parts: document analysis of 10 assessment of teaching plans and focus groups conducted with groups of faculty in two academic units selected to represent a range of experience assessing teaching using multiple measures.

3.1 Document Analysis

I chose to conduct document analysis because analyzing assessment of teaching plans would provide a broad base of data that focus group data could supplement (Bowen, 2009). Assessment of teaching plans also capture a significant amount of information, including academic

units' former assessment of teaching methods, the unit's approach to revision, and newly planned methods. Conducting document analysis allowed for the collection and analysis of a broad sample of academic units' plans. Of 14 academic units at the Pitt Main campus and three regional campuses, 13 assessment of teaching plans had been submitted by the time I began conducting this dissertation of practice. I chose to focus on analyzing the plans from academic units at Pitt Main. Although regional campuses each submitted plans, regional campuses each contain their own sub-units and are not comparable to the academic units (e.g. schools) at the Pitt Main campus. As a staff member who primarily serves the main campus, I would also have a limited ability to implement changes based on the results of this study at regional campuses.

I defined comprehensiveness by consulting assessment of teaching literature to identify the characteristics of effective assessment of teaching. Several common characteristics emerged across sources (Arreola, 2007; AAUP, 1975; Berk, 2006), which were used to establish rubric performance criteria (see Table 1: Characteristics of Effective Assessment of Teaching with Aligning Performance Criteria).

Table 1 Characteristics of Effective Assessment of Teaching with Aligning Performance Criteria

| Characteristics of effective assessment of teaching | Performance criteria |
|--|---|
| Academic units must create a discipline- and context-specific definition of teaching effectiveness against which to assess teaching. | Context-specific definition of teaching effectiveness |
| Faculty must be involved in the design of any systems used to assess their performance. Processes, criteria, and tools for evaluating teaching must be clearly communicated to and periodically reviewed and updated by faculty. | Faculty involvement |
| Methods of assessment should be varied and collect and triangulate evidence from different sources, including students, peers, review of artifacts, and self-assessment. Methods should align with the unit's definition of teaching effectiveness and account for differences in faculty rank, appointment, and course types and methods of delivery. | Multiple, varied sources of evidence for summative assessment |
| | Tools align with unit's definition of teaching effectiveness |
| | Flexibility to account for differences in rank/appointment, courses, methods of delivery, and/or faculty teaching style |
| Faculty evaluation and professional development should be designed to work together. If teaching is to be assessed, faculty need opportunities for formative assessment, feedback, and resources that will allow them to improve their teaching. | Addresses formative assessment and connects to faculty development |
| Assessment should be systematic and performed by qualified, well-prepared reviewers. | Reviewers and reviewer preparation |
| Assessments must product data that can be used for decision-making. | Use of results for continuous improvement of teaching |

To improve content validity, several experts from the Teaching Center, including the associate provost, the director of Teaching Support, and a manager in the Office of Measurement and Evaluation of Teaching, provided feedback on the performance criteria.

I created three performance levels for criteria: information missing or criterion not met, developing, and accomplished. Levels of performance are differentiated by completeness and/or quality. For context-specific definition of teaching, for example:

- At the level of “information missing or criterion not met,” the plan is missing a definition for teaching effectiveness.
- At the level of “developing,” the plan contains a definition, but the definition is vague or lists characteristics of teaching effectiveness that cannot be measured and/or characteristics that are unrelated to teaching.
- At the level of “accomplished,” the plan contains a context-specific definition that includes specific, measurable characteristics of teaching effectiveness.

In addition to these eight performance criteria, I also used plans to determine academic units’ approaches to developing their plans: (a) top-down, meaning administrators primarily made decisions and led the change process; or (b) middle-out, meaning groups comprised of mid-level unit leaders like associate deans and department chairs collaborated with faculty committee leaders and members to champion and lead the change process (see Appendix A: Document Analysis Rubric).

I established the standard for comprehensiveness as plans meeting all eight rubric criteria at the level of “developing” or higher. The eight performance criteria are all equally important, research-informed best practices. Omitting any criterion would make a plan less than comprehensive. However, it would be unreasonable to expect that academic units would achieve the “accomplished” performance level because plans represented concise first drafts of what was intended to be an iterative improvement process. In this early phase of the planning process, comprehensive plans should address all criteria, but may fall short of achieving the “accomplished” level.

I applied the rubric to the 10 academic unit plans. I replaced the descriptors of levels of performance with numbers (0-2 for the levels “information missing or criterion not met” –

“accomplished”) to determine plans’ overall scores. I recorded scores in Airtable, an online database platform. To ensure the credibility of ratings, my dissertation advisor acted as a second rater on two plans. The two of us agreed on 13 of 16 ratings and discussed discrepancies between ratings until we reached agreement.

3.2 Focus Groups

I conducted focus groups with participants from two academic units. Units were purposefully selected to represent differences in approach and experience with conducting multi-method assessment of teaching. One unit had been conducting assessment of teaching using multiple measures for over five years and had taken a top-down approach. The other began the process of selecting measures beyond teaching surveys in spring 2021 and had taken a middle-out approach. Focus groups consisted of six to seven volunteers from the groups of faculty members who had been involved in revising assessment of teaching and assessment plan creation. The majority of participants in both groups were women; each group had a single male participant. All participants in both groups were white. In the group from the unit that had taken a middle-out approach, participants held the following titles and ranks: associate dean, vice chair, program director or co-director, professor, and associate professor. In the group that took the top-down approach, participants had the titles and ranks of associate dean, chair, director, coordinator, professor, associate professor, assistant professor, and instructor. Participants in both groups had multiple titles and roles within their units. With the exception of one participant in the top-down group, all participants were full-time, tenured or tenure-track faculty. I am not listing participants’ disciplines to protect confidentiality.

I developed eight open-ended, semi-structured questions (see Appendix B: Focus Group Protocol and Interview Questions), which were slightly modified to reflect differences in the two groups' levels of experience conducting comprehensive assessment of teaching. My dissertation advisor reviewed my questions and focus group protocol (see Appendix B) to examine and offer feedback to improve content validity. Questions addressed:

- The unit's approach and process for revising assessment of teaching and creating plans;
- The unit's goals for revising assessment of teaching;
- The institutional factors, defined as structures, processes, policies, aspects of institutional culture, and/or resources that participants perceived to have affected plan creation;
- How assessment of teaching had changed from before and after implementing comprehensive assessment of teaching practices;
- What participants would like me and institutional leaders to know about assessment of teaching.

I collected the data during two 1 – 1.5-hour sessions via Zoom using an adapted Zoom focus group protocol (Dos Santos Marques, et al., 2020). Sessions were confidential but not anonymous and were conducted as private Zoom meetings. Because the samples were small, participants had extensive experience collaborating with one another, and questions were not sensitive in nature, I asked participants to keep their Zoom display names and cameras on throughout the session to encourage open dialogue. Participants gave verbal consent to be recorded for the purpose of transcribing their responses. Zoom generated an auto-transcript, which I revised to remove participant identifying information and to correct auto-transcription errors. I performed member checking during each session to determine that I understood participants' responses accurately and fully.

I coded focus group data inductively, using a combination of coding for patterns (Saldana, 2016) and constant comparison coding (Kolb, 2012) in NVivo, a qualitative data analysis software tool. That process consisted of considering each data set individually, then in comparison to the other, and iteratively revising codes and themes. I coded inductively because, although the review of literature on organizational change and assessment of teaching in higher education suggested potential codes, I wanted to remain open to the possibility that new information might emerge from the data. Examining patterns, similarities, and differences within and across cases was most appropriate for the overarching methodology, a comparative case study. In addition to expert review of the focus group protocol and member checking, I used thick description and direct quotations from participants. These strategies are particularly pertinent in improvement studies because they provide information about context which can inform transferability judgement.

4.0 Results

4.1 Analysis of Assessment of Teaching Plans

Of the 10 plans analyzed, four units took middle-out approaches, three took top-down approaches, and in three cases, the unit's approach was unclear based on the plan document. None of the plans met all eight criteria for at least the level of "developing," meaning that none of the plans were comprehensive. However, there was significant variation in plan comprehensiveness (see Table 2: Assessment of Teaching Plan Document Analysis Ratings).

Table 2 Assessment of Teaching Plan Document Analysis Ratings by Performance Criteria

| Approach/ Plan code* | Performance criteria | | | | | | | | Total score |
|-------------------------|-----------------------------|---------------------|------------------------------|------------|-----------------------------|------------------------------|-------------|----------------|-------------|
| | Context-specific definition | Faculty involvement | Multiple sources of evidence | Formative | Tools align with definition | Reviewers and reviewer prep. | Flexibility | Use of results | |
| Middle-out 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 2 | 15 |
| Middle-out 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 2 | 14 |
| Middle-out 3 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 14 |
| Middle-out 4 | 1 | 2 | 2 | 2 | 0 | 1 | 2 | 2 | 12 |
| Top-down 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 9 |
| Top-down 2 | 0 | 1 | 2 | 1 | 0 | 1 | 1 | 1 | 7 |
| Top-down 3 | 0 | 1 | 2 | 0 | 0 | 1 | 1 | 1 | 6 |
| Unclear 1 | 2 | 0 | 1 | 2 | 2 | 1 | 2 | 2 | 12 |
| Unclear 2 | 1 | 1 | 2 | 2 | 0 | 1 | 0 | 1 | 8 |
| Unclear 3 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 2 |
| Totals by criterion | 11 | 12 | 18 | 14 | 9 | 12 | 10 | 13 | |
| Averages by criterion | 1.1 | 1.2 | 1.8 | 1.4 | 0.9 | 1.2 | 1 | 1.3 | |

Note. A rating of 0 indicates that the criterion was missing or not met. A rating of 1 or 2 indicates that criterion was rated at the level of “developing” or “accomplished,” respectively (see Appendix A).

*I have chosen not to identify the disciplines of plans or to identify the units that participated in focus groups to preserve confidentiality.

Upon analysis, plan comprehensiveness seemed to differ based on the academic unit’s approach and how clearly the unit understood the expectations for creating assessment of teaching

plans. Although these two factors seem to account for differences in plan ratings, one anomaly did emerge: academic units tended to score low in flexibility, meaning flexibility of assessments to account for differences in faculty rank/appointment, courses, methods of delivery, and/or teaching style. Ratings for flexibility did not vary significantly according to approach or how clearly the academic units understood expectations for plan creation. Low ratings in flexibility may speak to differences in academic units' assessment of teaching goals.

4.1.1 Approach

Analysis of assessment of teaching plans showed that units that took middle-out approaches tended to have more comprehensive plans. In total, schools that took middle-out approaches scored at the “accomplished” level in 78% of all criteria, whereas schools that took a top-down approach scored at the accomplished level in 17% of all criteria. Schools that took middle-out approaches may have created more comprehensive plans because approach might reflect more faculty buy-in and the degree to which assessment and improvement of teaching are valued in the school. At minimum, the mid-level unit leaders and faculty who acted as champions of assessment of teaching in the unit valued the measurement and/or continuous improvement of teaching enough to spend time discussing teaching, devising new and improved strategies to assess teaching and apply assessment data, and contributing to assessment of teaching plans. Several patterns based on units' approaches emerged in plan criteria.

4.1.1.1 More Faculty Involvement in Middle-Out Plans

All units that took a middle-out approach indicated in their plans that multiple faculty members had a high degree of involvement throughout the plan creation process. Although units

that took a top-down approach all involved faculty in assessment plan creation to some degree, none of three top-down units engaged multiple faculty members to a substantial degree throughout the process. More faculty engagement suggests that units that took middle-out approaches were likelier to have incorporated faculty perspectives and expertise into the assessment of teaching plans they created.

4.1.1.2 More Context-Specific Definitions and Aligning Tools in Middle-Out Plans

Developing a context-specific definition of teaching effectiveness would have required discussion to identify and define characteristics of good teaching, which is perhaps why units that took middle-out approaches included definitions while units that took top-down often did not. All units that took a middle-out approach included a context-specific definition of teaching and three of four of these units included assessment tools aligning with the unit's definition. In contrast, only one of three units that took a top-down approach defined teaching effectiveness and addressed developing assessment tools to align with the definition. In that instance, the definition of teaching effectiveness was vague.

4.1.1.3 Use of Results for Improvement Beyond Individual Faculty Evaluations in Middle-Out Plans

All the middle-out plans and two of three top-down plans listed and described at least one formative assessment measure that the unit encourages faculty to use to inform improvements to teaching. Three of four middle-out plans versus one of three top-down plans identified multiple types of formative assessment that would be available to faculty in the unit and explicitly described how formative assessment connected to faculty teaching development efforts. All but one plan described how assessment of teaching would inform unit-level improvement of teaching, but only

four, including three middle-out plans and one plan for which the unit's approach was unclear, offered detailed descriptions including information about who in the unit will have access to and review data, how results will be communicated and shared, and the process for translating results into recommendations for improvement. This may reflect how faculty in the unit value the use of assessment data to improve teaching at the individual and unit levels, beyond using data for evaluative purposes. Faculty may also prefer to articulate who reviews and shares data because of concerns about confidentiality. Leaders may have offered fewer details in top-down plans because, in many cases already, they already have access to faculty evaluation data or presume that faculty in the unit have more knowledge of how data is reviewed and shared than they actually do.

4.1.1.4 More Middle-Out Plans Addressed the Need for Reviewer Preparation

Almost every plan identified who (e.g. department chair, dean, faculty committee) would review various sources of data, but only three plans, all from units that took middle-out approaches, included information about providing reviewers with training to conduct assessments and analyze and interpret results. Unit leaders like department chairs and deans are often responsible for reviewing summative assessment of teaching results to make hiring and promotion decisions. Therefore, units that took top-down approaches to assessment planning may be subject to unit leaders' assumptions of their own expertise and less likely to recognize the need to provide reviewers with training to ensure the consistency and equity of assessments. In units that took middle-out approaches, many of the faculty involved in assessment of teaching plan creation would not be involved in summative assessment data review but would be affected by the results of summative assessments, which may have made those units likelier to recognize the need for and value in reviewer training.

4.1.2 Understanding of Expectations

The degree to which units understood assessment of teaching best practices and expectations for creating plans may have also affected plan comprehensiveness. ACIE's recommendations for plan creation emphasized the need for multiple sources of evidence of teaching effectiveness including a peer review, student feedback, self-assessment, and a review of teaching artifacts. Likely because of the focus on this aspect of plans, all units included multiple measures of teaching effectiveness and eight of 10 units included all four types of measure. Academic units that scored a total of 12 or above out of 16 in terms of comprehensiveness of assessment of teaching plans met with me and, in some cases, the survey assessment manager of OMET to consult about assessment of teaching and plan preparation. This could indicate that these schools were more invested in this process or that they understood the expectations for plan creation more after meeting with Teaching Center staff. One of the least comprehensive plans was submitted prior to an extension of the assessment of teaching plan deadline, before plan guidelines and resources to aid in plan creation were communicated by the Teaching Center and the Office of the Provost. The lack of comprehensiveness of this plan was almost certainly due to a lack of understanding of what to address or include in the plan.

4.1.3 Flexibility

In criterion performance, units scored low (average score of one) in the flexibility of their plans, or how plan accounts for differences in faculty rank/appointment, methods of delivery, and/or faculty teaching styles and allows faculty to curate evidence of teaching effectiveness. Flexibility was also the only criterion for which units that took middle-out and top-down

approaches were rated similarly (though middle-out plans averaged a slightly higher average rating of 1.25). Focus groups, which were conducted to supplement and add rich faculty perception data to document analysis data, suggested a potential reason for this. One of the focus groups identified creating consistent standards for assessment of teaching across diverse departments as a primary goal during assessment planning. The other also communicated that they have a wide range of diverse programs within the unit but, rather than aiming for consistency, the unit preferred to tailor assessments to the needs of specific programs. Units attempting to establish consistency may have designed assessment of teaching to be less flexible, whereas units that prefer to customize assessments to the needs of departments or programs may have planned for differentiation.

4.2 Focus Groups

Constant comparison coding of focus group data revealed themes related to units' assessment of teaching revision goals, institutional factors that helped and hindered assessment of teaching planning, how revisions will or have changed assessment of teaching, and factors that are or will be necessary for assessment of teaching success. Shared patterns in codes across groups emerged from the analysis and comparison of data. Each group's timeline and process for revising assessment of teaching is presented to provide context for focus group data. In addition to shared codes, some codes were unique for each case, likely due to units' different reasons for beginning to revise assessment of teaching, approaches, and experience conducting assessment of teaching beyond the use of teaching survey data. Tables present an overview of codes, definitions, number of mentions by each group, and examples.

4.2.1 Units' Processes for Revising Assessment of Teaching

4.2.1.1 Middle-Out Group

One group described its unit's middle-out approach to revising assessment of teaching, which began in 2019. Faculty members and mid-level unit leaders interested in the continuous improvement of teaching formed a task force with representatives from each department. The task force instigated and sustained discussions about teaching effectiveness, acting as champions for teaching improvement work with support from the unit's leaders. The task force's initial goal was to define characteristics of effective teaching that could be applied across departments that are diverse in size and discipline. The task force ultimately passed the definition to an educational improvement committee. With support from the Assessment of Teaching Initiative, the committee initially focused on revising the unit's teaching survey questions, first to better align with the unit's definition for teaching effectiveness, then by using feedback from faculty and groups of undergraduate and graduate students. When the Provost released the ACIE recommendations for assessment of teaching and the charge to develop an assessment of teaching plan by the end of the spring 2021 semester, the committee responded by conducting a survey to determine what assessment of teaching methods were currently being used by departments within the unit. Survey results revealed that practices were inconsistent across departments and primarily consisted of conducting teaching surveys. The committee collaborated to create the unit's assessment of teaching plan and, beginning in the summer of 2021, began developing tools and processes for faculty self and peer assessment that align with the unit's definition of teaching effectiveness. The unit will pilot the new tools and processes in 2021-2022.

4.2.1.2 Top-Down Group

The other group's unit adopted a top-down approach to revising assessment of teaching as part of a larger effort to meet specialized accreditation standards, beginning in 2016. Unit leaders formulated policies and practices related to changes with some input from the unit's curriculum committees. Programs within the unit vary significantly in size, discipline, and primary instructional methods. The unit developed an integrated approach to assess teaching, student learning outcomes, and curriculum, but each program has tailored processes and tools to their own needs. The unit recently established an office which provides faculty development support and assists programs with assessment of teaching. Because the unit has been conducting assessment of teaching using multiple measures for longer than other units at Pitt, its assessment of teaching plan involved documenting the work the unit has already done and describing planned revisions rather than developing entirely new policies, processes, and tools. Although participants of this group identified their approach as top-down, they also described how unit leaders have increasingly communicated with and involved faculty more in teaching assessments, which suggests that the unit may be shifting to a more middle-out approach over time.

4.2.2 Goals for Revising Assessment of Teaching

Both groups described continuous improvement, meaning the desire to build processes and structures that support the ongoing improvement of teaching and learning as a primary goal for revising assessment of teaching (see Table 3: Goals Theme). The groups specified that the continuous improvement rather than compliance with standards or policies external to the unit should be the primary purpose of assessment, with one participant from the top-down group explicating that,

part of the reason we wanted to revise our processes is I don't want it to be like a checkmark, like, oh, we looked at this course and it's good for three years. Like, I really want to look at courses, look at student feedback, and what changes can we make to produce an impact-you know, improve outcomes.

Although both groups expressed the desire for continuous improvement, the middle-out group focused on improvement of teaching whereas the top-down group referred to overall program improvement, including teaching. The top-down group's unit uses an integrated model for assessing teaching, learning, and curriculum, which is likely why that group's participants spoke about improvement more broadly.

Some patterns in codes emerged only or primarily from the middle-out group's discussion (see Table 3), likely due to differences in the two units' phases of assessment revision. The most prevalent pattern from the middle-out group's discussion of goals was the desire to adopt additional assessment methods beyond teaching surveys. Comments indicated that participants perceived conducting teaching surveys alone to be problematic because student feedback only offers a limited perspective on teaching effectiveness and teaching survey data can be influenced by student biases. Participants expressed that they want to use additional assessment methods that would generate more robust data from multiple sources. Until recently, teaching surveys were often the only measure used by departments in the middle-out group's unit, whereas the top-down group's unit has used many measures, including multiple sources of student feedback, for years. The middle-out group also discussed wanting to create faculty development resources and support, particularly for new faculty, as a key goal. The top-down unit already has faculty support infrastructure in place, including an office devoted to faculty development and, for the largest program in the unit, processes for conducting formative assessment and improvement of teaching

in collaboration with the Teaching Center. The middle-out unit only recently hired a faculty member who will lead faculty development efforts. The unit’s teaching support staff largely focus on the instructional design of online courses. The middle-out group’s unique goals likely reflect that their unit is in a much earlier phase of assessment of teaching revision and has had less time to develop resources to help faculty assess and improve their teaching.

Table 3 Goals Theme

| Code and definition | Middle-out mentions and examples | Top-down mentions and examples |
|---|--|--|
| <p>Moving beyond teaching surveys The desire to adopt additional assessment methods, gather evidence from other sources like peers, produce richer assessment data. Includes mentions of biases associated with teaching survey.</p> | <p>Mentions: 14 “But you know our goal was, again, to look beyond the OMET surveys.” “I think using the peer review process versus OMET will give us different types of data, perhaps.”</p> | <p>Mentions: 1 “We have a very small number of students participating [in the OMETs] which skews the results toward low numbers. It’s known that the people who are unhappy, they will be more willing to do the evaluations.”</p> |
| <p>Continuous improvement The desire to build processes and structures that support the ongoing improvement of teaching and learning; not approach assessment from a compliance perspective</p> | <p>Mentions: 9 “Let’s hone in on maybe one or two specifics about teaching. What did you learn that worked well this year in terms of teaching? What did you find that didn’t work well? What are your plans to address that, trying to close the loop, or at least create a loop? That there’s not just accountability, but there’s growth. There’s development. There’s some level of expectation. And I can see that we’re moving in that direction...And I think that’s what we want.”</p> | <p>Mentions: 4 “I think part of the reason we wanted to revise our processes is I don’t want it to be like a checkmark, like, oh, we looked at this course and it’s good for three years. Like, I really want to look at courses, look at student feedback, and what changes can we make to produce an impact-you know, improve outcomes.”</p> |
| <p>Consistent assessment of teaching standards The academic unit lacks and wants consistent standards and guidelines for assessment of teaching across departments.</p> | <p>Mentions: 7 “I think there was motivation to have some guidelines that would go across departments, because you know from the survey [of prior assessment of teaching practices in departments], it was definitely clear that it’s being done in different ways in different departments and there wasn’t really any guidelines kind of school-wide on how to approach it or whether it was even important to do other than the university requirement to do the OMETs.”</p> | <p>Mentions: 0</p> |

| | | |
|---|---|---|
| <p>Providing resources and support for faculty Desire to create processes, tools, and resources that support faculty development, particularly new faculty development</p> | <p>Mentions: 6 “So how do we kind of expand those [standards for teaching excellence] and help young faculty learn what teaching excellence is and built it into their repertoire, you know, as part of that process as well? So I think that was also a guiding force.”</p> | <p>Mentions: 1 “We wanted to make sure that we had a clear way for [faculty] to know what the resources were, what changes need to be made, and kind of what direction to move in.” (1)</p> |
| <p>Rewarding good teaching Want to create incentives to improve teaching; Tangibly value and support teaching, not just research</p> | <p>Mentions: 7 “I think faculty need to see that there is a direct link between excellence in teaching and whatever is related to your position as far as promotion and all other types of benefits that are associated with that. And seeing that link, I think, is a big motivator for people, whether it be a teaching award or whether it be a promotion or whether it be a raise.”</p> | <p>Mentions: 0</p> |

Note. Less frequent codes included: giving students a high-quality learning experience (4 mentions), clarify path to improvement (3 mentions), and sustainability (2 mentions)

The groups also commented on the diversity of departments and programs in their unit, which vary by number of students; numbers and ratios of research, teaching, and clinical faculty; discipline, and primary instructional methods. Although both groups mentioned this, discussion with the middle-out group demonstrated the unit’s interest in establishing consistent assessment of teaching standards across departments as another goal. Rather than using the same standards, the top-down group explained that assessment of teaching in their unit has been designed to vary by program needs and interests. The top-down group’s unit shares a common goal of wanting to improve teaching but does not use common standards or measures. As one participant stated, programs, “share goals and values, but the mechanism may be different or the focus may be different.” The middle-out group may prioritize consistency of standards because the unit lacks data on teaching effectiveness and hopes to establish a common baseline. The middle-out group’s unit also articulated a unit-level definition of teaching effectiveness, whereas the top-down group

did not. This would make it easier and potentially more desirable for the middle-out group's unit to establish school-wide standards aligning with the definition. Units' cultures may be another contributing factor. The degree to which departments or programs within a unit collaborate or are siloed could affect willingness to adopt common standards. Participants in the middle-out group described wanting to foster more cross-departmental collaboration and discussion, whereas participants in the top-down group specified that they could each only speak to the practices in their programs. Whether faculty and unit leaders value attaining a standard level of teaching quality as in the middle-out group or customizing assessment of teaching processes and tools to the potentially unique needs of a department or program as in the top-down group may account for differences in the two units' responses.

The desire to reward good teaching was discussed only by the middle-out group (see Table 3: Goals Theme). Again, this could be due to units' different cultures. The middle-out group's approach, which was led by faculty and mid-level leaders, indicates that a contingent of faculty in the unit value teaching and instructional improvement. Members of the middle-out group also remarked that research is valued more than teaching. One participant described how tenured faculty in particular prioritize research over teaching,

I think sometimes there's a gap. Sometimes, some people who are in tenure track positions feel like [their] main job is to get research dollars and that's what [they're] rewarded for. It *is* what they're rewarded for and they don't necessarily...it's not consistent that they put effort into their teaching.

If a group of faculty members in the unit valued teaching as much as they do research, it follows that they would want to create incentives to reward good teaching like research activity is rewarded. The top-down group may not have identified rewarding teaching as a goal because they

believe that faculty evaluation and incentives already align with unit values (whether that includes valuing teaching or research more). Having conducted multi-method assessment of teaching for a longer period of time, the top-down group's unit may have already made changes to evaluation and incentives. The top-down group's leader-led approach may have also played a role. If faculty in the unit feel little ownership or control over assessment of teaching policies and processes, they may feel that they are unable to change how aspects of faculty work are rewarded.

4.2.3 Institutional Factors that Affected Assessment of Teaching Planning and Factors

Necessary for Success

Focus groups identified institutional factors, or structures, processes, policies, characteristics of unit culture, or resources that aided in or impaired assessment of teaching revision planning. Patterns in institutional factors were somewhat consistent between the two groups. Both described how unit culture and access or lack of resources affected assessment of teaching planning. The code most frequently mentioned as an institutional factor that aided in assessment of teaching planning across the groups was access to resources (see Table 4: Institutional Factors that Helped Assessment of Teaching Theme). All the resources that the groups identified as helpful were departments within the institution but external to their own units that provided support in the planning or implementation of assessment of teaching. Groups described Teaching Center services as helpful, likely in part because they were speaking to me, a Teaching Center staff member. In addition to addressing assessment of teaching planning and teaching improvement support, participants acknowledged that the Teaching Center's Office of Measurement and Evaluation, which deploys, collects results of, and reports on teaching surveys, removes some of the burden of collecting student feedback from faculty. Participants in the top-

down group also commented that the Center for Doctoral and Postdoctoral Career Development became an important resource to the unit by sharing assessment methods and tools.

Table 4 Institutional Factors that Helped Assessment of Teaching Theme

| Code and definition | Middle-out mentions and examples | Top-down mentions and examples |
|--|--|--|
| <p>Access to resources Access to resources aided in assessment of teaching planning. Includes mentions of the Center for Teaching and Learning, the Center for Doctoral and Postdoctoral Career Development, and teaching surveys.</p> | <p>Mentions: 2 “The Task Force did seek out the resource from the Teaching and Learning Center. So there was a person that came to multiple sessions and talked and worked and gave some background. So that was a plus in terms of resources there.”</p> | <p>Mentions: 13 “And we found a lot of interesting information in, actually, the office across the street, [The Center for Doctoral and Postdoctoral Career Development]. They have a very sophisticated program of assessment of postdocs and we think that some of this can be translated into our program.”</p> |
| <p>Culture The unit shares a commitment to assessing and improving teaching. Includes mentions of desire for increased consistency of assessment of teaching across departments and faculty collaboration on assessment of teaching work.</p> | <p>Mentions: 10 “I just want to acknowledge that this has been a group effort, certainly, in [this school].”</p> | <p>Mentions: 3 “[Programs] share goals and values, but the mechanism may be different or the focus may be different.”</p> |
| <p>Drive from the top Clear, consistent messaging from leaders that teaching and assessment of teaching are valuable and alignment with a strategic plan that emphasizes teaching effectiveness has helped drive assessment of teaching planning and revision.</p> | <p>Mentions: 9 “We had a change in administration. So we did have a new dean and we would hear from him saying, ‘We really do have to come up with ways to reward excellence in teaching.’ And he would say that many times, I think, over his beginning tenure, and even now. So I think that added to it, as well as some of the changes that we’ve seen across the university with a strategic plan and so forth.”</p> | <p>Mentions: 1 “It was led really by the curriculum committee leaders at the time. I was chair and a colleague of ours who’s no longer here was co-chair and he was really the leader of developing that policy and, you know, tried to involve the curriculum committee. But in terms of actually getting things together and getting literature to support not just using one form of evaluation and assessment, I would say it was driven more by him and me.”</p> |

Note. Less frequent codes included: other types of student feedback (5 mentions), generating buy-in (3 mentions), and driven by accreditation (3 mentions)

Lack of resources emerged as an institutional factor that hindered assessment of teaching planning (see Table 5: Institutional Factors that Hindered Assessment of Teaching Theme). Participants cited lack of faculty time to engage in assessment of teaching work as one key issue. One participant expressed concern that assessment of teaching work might fall disproportionately to teaching faculty in the unit, stating, “it’s the teaching faculty that might end up having this time-intensive-I don’t want to say burden-but it’s more layers of work.” Both groups mentioned peer reviews specifically as a time-consuming assessment method. Beyond faculty’s limited time, participants in the top-down group also expressed that they were unsure of whether they were aware of all the support resources available or where to go for assessment of teaching assistance. As one participant explained, “I feel like sometimes people just don’t know where to go to get help. And it’s like we know the Center for Teaching and Learning exists. Like, I know I could email you and reach out, but I’m sure you’re busy, and so it’s just a who to approach and where to go to sort of facilitate that process.” Some participants in the top-down group further specified that the perceived lack of centralized resources to support assessment of less traditional instructional modes like mentoring hindered assessment of teaching. Although the Assessment of Teaching Initiative does curate resources tailored to faculty’s requests, some participants in the top-down group were unaware of this service.

Another shared pattern in institutional factors that affected assessment of teaching across the two groups’ responses was unit culture. The units both described cultures of commitment to assessing and improving teaching, which aided in assessment of teaching planning, while the diversity of departments and programs in their units sometimes made assessment of teaching planning more challenging (See Tables 4 and 5). Members of the middle-out group also identified the prioritization of research over teaching as an aspect of unit culture that hindered assessment of

teaching planning. One participant questioned whether the unit would be able to make the cultural shift necessary to implement changes:

Will we be able to change with it and not give our focus to [teaching surveys] and give the same value, which I think maybe [other participant] was talking about, to the peer review parts? 'Cause I don't know that we... We tend to, you know... How do we teach kind of like an old dog new tricks? And just say, yeah, we are going to focus more on that?

The top-down group did not mention the need for cultural shift, likely because the unit has been conducting assessment of teaching using multiple measures for longer. At this point in the process, the top-down group's unit is revising assessment of teaching rather than instituting completely new processes.

Table 5 Institutional Factors that Hindered Assessment of Teaching Theme

| Code and definition | Middle-out mentions and examples | Top-down mentions and examples |
|--|--|--|
| <p>Culture Aspects of unit culture including the diversity of disciplines and approaches to assessment of teaching within the unit and the fact that research is valued more highly than teaching made assessment of teaching planning and revision more challenging. Includes mentions of the difficulty of shifting unit culture.</p> | <p>Mentions: 16 “Will we be able to change with it and not give our focus to [teaching surveys] and give the same value, which I think maybe [other participant] was talking about, to the peer review parts? ‘Cause I don’t know that we... We tend to, you know... How do we teaching kind of like an old dog new tricks? And just say, yeah, we are going to focus more on that?” “One of our things too, is our departments are, across the board, the school, it’s so different. We have very small departments with a few faculty, then we have some really large ones. Or we do have a small number of teaching faculty, but a lot of research folks who don’t really do a lot of teaching.”</p> | <p>Mentions: 3 “[Programs] share goals and values, but the mechanism may be different or the focus may be different.”</p> |
| <p>Lack of resources Lack of resources like time or lack of awareness of resources or how to access them hindered assessment of teaching planning. Includes mentions of lack of centralized support for assessment of mentoring.</p> | <p>Mentions: 7 “I think the concerns are about the time intensiveness of it.”</p> | <p>Mentions: 9 “We’ve got all of these [faculty experts], but everybody’s so busy that, again, trying to figure out how to do some peer review without adding, you know, too much of a burden. So resources influences it, but it’s sort of a lack of resources.” “I feel like part of it now, especially for newer faculty-I just did a faculty review a couple of weeks ago and I feel like faculty struggled to know where to reach out to.”</p> |
| <p>Challenge of generating buy-in Generating widespread support of assessment and improvement of teaching can be challenging.</p> | <p>Mentions: 6 “I think it has to be as important for tenure track people too. I think sometimes, there’s a gap. Sometimes, some people who are in tenure track positions feel like, ‘My main job is to get research dollars and that’s what I’m rewarded for.’”</p> | <p>Mentions: 0</p> |

Note. Less frequent codes included: teaching surveys are less helpful than other forms of student feedback (5 mentions), anxiety-provoking (3 mentions)

When discussing which institutional factors would be necessary for assessment of teaching work to succeed, groups commented on feasibility and sustainability (see Table 6: Institutional

Factors that Will Be/Are Necessary for Assessment of Teaching Success Theme). A participant in the top-down group described having to revise assessment of teaching processes to address these issues:

So what we found is that there were some parts of the [assessment of teaching] policy that were not sustainable and that became really clear during the pandemic when everybody was working so hard just to kind of keep up and get the curriculum going and get their students graduated.

Participants emphasized that assessment of teaching processes must be designed to be manageable for faculty balancing competing priorities and heavy workloads.

Table 6 Institutional Factors that Will Be/Are Necessary for Assessment of Teaching Success Theme

| Code and definition | Middle-out mentions and examples | Top-down mentions and examples |
|--|--|--|
| <p>Drive from the top Clear, consistent messaging from leaders that teaching and assessment of teaching are valuable and alignment with a strategic plan that emphasizes teaching effectiveness is necessary to support assessment of teaching planning and revision.</p> | <p>Mentions: 9 “To me, one thing that’s going to be important, you know, in your role, [other participant], is that the message comes top-down to all the chairs that this needs to be included.”</p> | <p>Mentions: 0</p> |
| <p>Feasibility and sustainability New assessment of teaching processes must be designed to be feasible for and that they can be sustained by faculty balancing competing priorities.</p> | <p>Mentions: 4 “And there’s also a focus on making sure that it is manageable, right? That this is a feasible process that can feed in. And I think all of those things work together to create a process that people are going to want to engage in.”</p> | <p>Mentions: 2 “So what we found is that there were some parts of the policy that were not sustainable and that became really clear during the pandemic when everybody was working so hard just to kind of keep up and get the curriculum going and get their students graduated.”</p> |

Note. Less frequent codes included: transparency (4 mentions), accountability (3 mentions), flexibility (2 mentions)

As with goals, some patterns in codes related to institutional factors were unique or predominant in the middle-out group's discussion. The middle-out group described the importance of drive from the top, meaning clear, consistent messaging from leaders that teaching and assessment of teaching are valuable and align with unit strategic goals (see Tables 4 and 6). Drive from the top was identified as an institutional factor that advanced and was necessary for the success of assessment of teaching work. The group also mentioned generating buy-in from colleagues as a factor that made planning more challenging (See Table 5). For the top-down unit, support from leaders was a given. The top-down approach and need to meet accreditation standards may have also made generating faculty buy-in at least initially less important. The top-down unit may have also moved beyond the early phases of assessment of teaching revision when generating faculty buy-in would occur; the unit might already have cultivated buy-in.

4.2.4 How Revisions Will/Have Changed Assessment of Teaching

When discussing how revisions will or have changed assessment of teaching, patterns in groups' responses indicate that both groups have observed or expect positive changes (see Table 7: How Revisions Will/Have Changed Assessment of Teaching Theme). The groups felt that conducting additional measures of teaching effectiveness would be beneficial and that the quality and completeness of assessment of teaching data would improve. In the middle-out group's case, the unit had previously relied on teaching surveys as the primary means of assessing teaching. Participants in the middle-out group frequently mentioned that the peer review process that the unit was developing would allow faculty to gather valuable feedback from colleagues, not just students. Participants in the top-down group discussed developing processes and tools to assess mentoring for the first time, which would capture an important aspect of instruction that had not

previously been assessed. Having adopted integrated assessment of teaching, learning, and curriculum, participants in the top-down group also discussed the improvement of assessment data overall.

Each group reported that changes would increase the equity of assessment, although how participants defined equity varied. Members of the middle-out group cited the use of additional measures beyond potentially biased teaching survey data as an improvement. As one participant explained,

Even if you just talk about [teaching surveys], and I don't know how different departments use that, but we know there are inherent biases and things that are associated with that, so continuing to allow that to serve as the primary source of how we determine teaching effectiveness just didn't make sense.

Another participant in the middle-out group mentioned that reviewer training would make assessment of teaching more equitable, stating, "I just feel like the peer review process, which will include teaching peer reviewers, so we're going to do the training of reviewers to hopefully help with equity and decrease biases." Rather than reducing the potential of bias to affect assessment of teaching data, members of the top-down group commented on increasing the consistency of and faculty involvement in assessment of teaching. Several participants identified performing calibration exercises to ensure that assessments were being conducted in the same way as a means by which assessment had become more equitable.

Table 7 How Revisions Will/Have Changed Assessment of Teaching Theme

| Code and definition | Middle-out mentions and examples | Top-down mentions and examples |
|--|---|---|
| <p>Provide additional measures Revisions will create additional measures and data sources to assess teaching effectiveness beyond teaching surveys.</p> | <p>Mentions: 5 “I think the data points are totally new when you have a peer reviewer versus a student.”</p> | <p>Mentions: 4 “We are more interested in the mentoring part and three years ago, we didn’t have any [assessment of mentoring]. So now we started to develop it. So I guess over time, it changed from zero [methods of assessing mentoring] to three.”</p> |
| <p>Increase equity Revisions will make/have made assessment of teaching more equitable.</p> | <p>Mentions: 3 “I just feel like the peer review process, which will include teaching the peer reviewers-so we’re going to do the training of reviewers to hopefully help with equity and decrease biases.”</p> | <p>Mentions: 6 “We have not defined equity and we have not defined that these are the things we do for equity of assessment. However, as [another faculty member] mentioned, it’s more of a top-down process and it’s always been a top-down process. However, faculty have the opportunity to voice their opinions.”</p> |
| <p>Improve data Revisions will/have increase(d) the quality and completeness of assessment of teaching data.</p> | <p>Mentions: 3 “[Peer review] information should be so meaningful and give someone...it should be so much more rich, I guess, because it’s multidimensional...”</p> | <p>Mentions: 5 “Um, I think it’s improved for sure. I think it’s helpful to have...I love data, so I love like online tools that track data. For me, then I can look back and compare things. It’s also helpful as I feel like we made a lot of changes in the program and you want to know how those changes impact students.”</p> |

5.0 Learning and Next Steps

5.1 Discussion

The purpose of this dissertation of practice was twofold. The first aim was to determine whether one approach to revising assessment of teaching, top-down or middle-out, was more successful when it came to developing comprehensive assessment of teaching plans, defined as plans that met eight research-derived criteria for successful assessment of teaching at the level of “developing” or higher. Although none of the plans met the standard for comprehensiveness, plans created by academic units that took middle-out approaches were rated higher than plans of academic units that took top-down approaches. This may have been due to the higher faculty support and engagement required by a middle-out approach. Developing plans based on broader faculty input and expertise may have led units that took middle-out approaches to create more comprehensive plans.

The second purpose of this dissertation of practice was to examine the institutional factors that faculty perceived to have affected the assessment of teaching process to learn which factors drive and impede assessment of teaching revision planning. Focus group participants came from two units with different approaches to and experience with conducting assessment of teaching, which likely informed some differences in groups’ responses. Nevertheless, there were some commonalities across both groups. Both reported the desire for continuous improvement (of teaching or overall program quality) as a goal for revision of assessment of teaching. Groups identified unit culture and access or lack of access to resources as institutional factors that both helped and hindered assessment of teaching revision. They also indicated that assessment of

teaching processes must be feasible and sustainable. Both groups felt that the revisions they had outlined in their unit's assessment of teaching plans would generate more evidence of teaching effectiveness, create a more equitable assessment process, and provide faculty with more complete, robust assessment of teaching data.

Research on assessment of teaching best practices and organizational change in higher education suggested that several institutional factors, including unit culture, unit leaders' and faculty members' attitudes towards change, and access to resources could affect this type of institutional change. The results of the assessment of teaching plan document and focus group data analyses confirmed that these factors can act as drivers and barriers of this change process, but also suggested that differences in a unit's approach and level of experience conducting multiple measures of teaching effectiveness may affect the degree to which and how the faculty in the unit perceive these institutional factors. Assessment of teaching plans from units with middle-out approaches were rated higher for comprehensiveness, which aligns with recommendations to charge collaborative teams rather than leaders with leading strategic change efforts (Andrews et al., 2020; Eckel et al., 1999; Finkelstein et al., 2020; Kezar, 2018; Reinholz et al., 2017; Weaver et al., 2020). Taking a middle-out approach may have also led to the middle-out focus group participants reporting that support from unit leaders and cultivating buy-in from their colleagues drove revision of assessment of teaching and created challenges, respectively. The top-down group did not mention these factors, potentially because a top-down approach is leader-driven or because that group's unit had been conducting multiple measures of teaching effectiveness for longer and already had the support needed to advance the process. The middle-out group's emphasis on the importance of cultivating leader's and faculty's interest and support for teaching assessment and improvement revision echoed the literature on organizational change and assessment of teaching

(AAUP, 1975; Andrews et al., 2020; Arreola, 2007; AAUP, 1975; Boyce, 2003; Finkelstein et al., 2020; Kezar, 2018; Murphy et al., 2009; Storberg-Walker & Torraco, 2004; Weaver et al., 2020). Without investment from leaders, faculty might lack the authority to embed new assessment of teaching processes in formal faculty evaluations. Without support from faculty, leaders could implement changes, but faculty could treat assessment as a policy with which to comply rather than a continuous improvement process.

Both focus groups stated that unit culture and resources were the major institutional factors that affected assessment of teaching revision planning. While shared goals and values within the unit helped advance assessment of teaching revision, diversity of programs and disciplines was identified as an impediment. This is consistent with cultural theories of change in higher education, which propose that successfully enacting widespread, transformational change necessitates building a shared vision. (Boyce, 2003; Kezar, 2018). Theories of change also suggest that deep change can require shifting the beliefs and values of leaders and faculty (Boyce, 2003; Kezar, 2018). The middle-out group discussed the need to change unit culture and described how the unit's practice of rewarding research more than teaching improvement was a factor that hindered assessment of teaching planning. The middle-out group also identified rewarding teaching improvement as a goal of their unit's assessment of teaching revision work. Guides and studies on teaching assessment indicate that incentivizing and rewarding teaching by tying it to formal evaluation processes can encourage teaching improvement (Arreola, 2007; Bradforth, et al., 2015; Finkelstein et al., 2020; Weaver et al., 2020). In the top-down group's unit, which started revising assessment of teaching earlier, faculty evaluation processes may have already been changed to reflect revisions. It is also possible that, due to top-down decision-making structure, faculty in that unit may feel that they are unable to propose revisions to faculty evaluation processes. These may

be reasons why the topic of rewarding teaching did not come up in the top-down group's discussion.

When the focus groups discussed access to resources, some responses were consistent across groups. Both groups identified the Teaching Center as a helpful resource, although the top-down group was unaware of some Teaching Center services. Several studies on organizational change noted the importance of providing resources to support change processes (Bradforth et al., 2015; Eckel et al. 1999; Finkelstein et al., 2020). Research on assessment of teaching best practices suggested that institutions should share common resources campus-wide (Finkelstein et al., 2020) and that faculty development efforts should be explicitly linked to assessment of teaching (AAUP, 1975; Arreola, 2007). A recent article (Sacks et al., 2021) from the Professional and Organizational Development (POD) Network in Higher Education, the largest national organization of educational developers, also made the argument that Centers for Teaching and Learning could be leveraged to support accreditation and other quality assurance processes because Teaching Center staff often possess expert knowledge of improving teaching, curriculum, and programs. The Center for Teaching Excellence (CTE) at the University of Kansas, one of the institutions in the TEval NIC, provides a model for Teaching Center leadership in revision of assessment of teaching (Finkelstein et al., 2020; Weaver et al., 2020). The University of Kansas' CTE provides departments with assessment of teaching support by liaising and consulting with departmental project team members, curating examples and resources for departments, and facilitating a cross-department working group that meets several times a year to share their progress and learning (Finkelstein et al., 2020; Weaver et al., 2020). Taken with focus group participants' responses, this indicates that Teaching Centers can provide vital unit-level support of teaching assessment and improvement, beyond their more traditional individual instructor-level support. The top-down

group identified a lack of a centralized support resource for assessment of less traditional modes of teaching as an institutional factor that hindered their unit's assessment of teaching revision planning. Although Pitt has a centralized support unit, the Teaching Center, lack of awareness of faculty support services has been identified as a major barrier to teaching improvement efforts in past studies (Gillespie & Robertson, 2010; Halasek et al. 2020). The absence of resources for assessing less traditional types of teaching also represents a gap in Pitt's, and possibly other institutions', resources.

The focus groups also identified lack of faculty time as a barrier. Participants in both groups expressed concern about conducting more time-consuming assessments like peer observations. This concern may have contributed to the focus groups identifying feasibility and sustainability as factors necessary for developing successful assessment of teaching processes. Research on change management in higher education repeatedly points to lack of faculty time as a significant obstacle and source of resistance in implementing change initiatives (Bradforth et al., 2015; Brownell & Tanner, 2012; Eckel et al., 1999, Kezar, 2018; McMurtrie, 2019). Restructuring time and priorities, spreading out tasks amongst a larger group of participants, and embedding change activities within existing events like meetings can mitigate this issue (Eckel et al., 1999; Weaver et al., 2020). Both focus groups discussed considering these strategies to make more time-consuming assessments more sustainable for faculty.

Patterns across units that emerged from document and focus group data analysis verified findings from higher education organizational change and assessment of teaching research that indicated that revising assessment of teaching constitutes a multi-level organizational change. Academic units' approaches and broad categories of institutional factors, particularly access to resources, and unit culture affect units' assessment of teaching revision planning. However,

differences in patterns from the two focus groups' discussions also illustrate the importance of examining a change within its unit-level context, not just the institutional context (Andrews et al, 2020; Kezar, 2018; Weaver et al., 2020). The results from this dissertation of practice suggest that while research on organizational change and assessment of teaching can provide a general understanding of the potential drivers and barriers for revising assessment of teaching, examining the specific context of a unit, including the unit's approach for managing change, culture, attitudes of leaders and faculty, goals, process, and timeline is also necessary. Facilitating revising assessment and improvement of teaching requires supporting multiple levels of change with attention to aspects of units' contexts.

5.2 Next Steps and Implications

5.2.1 At Pitt

As academic units at Pitt move forward with implementing their assessment of teaching plans, the results of this study suggest that units should support first-order change by building the basic infrastructure, capacity, and resources and promote second-order, cultural shift that might be necessary for faculty to prioritize teaching more. Table 8 lists stakeholder groups and summarizes support strategies, which are detailed below. Although strategies for support are loosely categorized, categories are nondiscrete. For example, some strategies that build infrastructure, capacity, and resources may also contribute to supporting cultural change.

Table 8 Stakeholder Groups and Strategies for Support

| Stakeholder group | Strategies for support | |
|-----------------------|--|--|
| | Building infrastructure, capacity, resources | Supporting cultural change |
| Faculty | Form collaborative, representative teams and distribute assessment of teaching tasks. Consider and assess feasibility, sustainability, impact, faculty perceptions of equity and data quality and make adjustments as needed. Adopt an integrated approach to assessment to improve efficiency and data use. | If not already done, articulate unit-specific definition(s) of teaching effectiveness. Work to align processes and methods with unit culture. Champion the cause amongst colleagues. Consider how graduate students might be engaged in this process. |
| Unit leaders | Allocate time, resources, and attention to assessment of teaching work. Embed assessment of teaching within existing unit events and meetings | Create incentives and rewards for participating in assessment of teaching and improving teaching, including embedding assessment of teaching in formal evaluation processes. Use and communicate the use of results |
| Administration | Work with unit leaders, faculty, and the Teaching Center to determine and communicate next steps and how results will be tracked and used. | Increase the visibility of assessment of teaching work. Voice support for assessment of teaching work. |
| Teaching Center staff | Provide generalized and unit-specific support of assessment of teaching. Assist with or facilitate unit assessment of teaching meetings. | Celebrate assessment of teaching successes. |

Middle-out approaches produced more comprehensive assessment of teaching plans. Units have already established an initial change approach, but approaches could evolve over time as evidenced by the top-down groups comments about increasing faculty involvement. For those that initially took top-down approaches, engaging faculty more in collective decision-making and developing assessment of teaching processes, tools, and policies might improve some aspects of assessment of teaching plans that were rated lower, particularly faculty involvement. More faculty involvement might also help cultivate the buy-in necessary for faculty to engage in assessment of teaching in a meaningful way. A team-based approach also allows work to be distributed amongst

members of a group rather than left to a few individuals (Eckel et al., 1999; Finkelstein et al., 2020; Kezar, 2018; Reinholz et al., 2017; Weaver et al., 2020). Creating collaborative teams with members who are passionate about improving teaching and are representative of the unit would ensure that various voices and perspectives are heard as assessment of teaching plans are implemented (Eckel et al., 1999; Finkelstein et al., 2020; Kezar, 2018; Reinholz et al., 2017; Weaver et al., 2020). These groups should include instructors of various ranks and appointments, including part-time instructors and graduate students. Part-time instructors and graduate students are often overlooked as stakeholders in teaching improvement efforts, even though they are subject to assessment of teaching and participate in teaching improvement. Because participating in an institutional change initiative would fall outside of their typical duties, these stakeholders should also be compensated for their valuable input.

One of the first tasks for unit teams that have not already done so would be discuss and develop a shared vision of teaching effectiveness and continuous improvement of teaching. Developing common values can drive assessment of teaching work and will allow units to compose aligning criteria and standards for teaching effectiveness. Units will need to determine whether their faculty prioritize increasing the consistency of teaching quality or customizing definitions and assessments of teaching by department or program. Creating more opportunities for collaborative faculty teams to engage in, discuss, and lead assessment of teaching can support both first and second-order change because teams can alter structures, processes, and the distribution of work, but they can also serve as a locus for cultural change. Team members can act as ambassadors for change to their colleagues.

Units, particularly those that took or are gradually moving toward middle-out approaches, need explicit support from unit and institutional leaders to implement changes. Leaders can show

support by allocating resources, time, and attention to assessment of teaching work (Andrews et al., 2020; Eckel et al., 1999; Kezar, 2018; Weaver et al., 2020). For faculty participating in assessment of teaching, unit leaders can demonstrate support by making adjustments to existing events or processes to give faculty time to engage in assessment of teaching. These might include reserving time for assessment of teaching work within existing unit meetings or giving faculty course releases, protected time, or recognition of assessment of teaching as service work. Units could also improve the efficiency and effectiveness of assessment by considering measures that provide data on multiple aspects of the student learning experience. Adopting an integrated approach to assessing teaching, curriculum, and learning outcomes may save faculty time and allow units to examine overall program quality.

Although carving out time removes one barrier for faculty who want to participate in assessment of teaching, providing time alone does not promote cultural change. Unit leaders must also incentivize and reward teaching assessment and improvement. Units' promotion and tenure policies typically address but may not adequately reward teaching effectiveness. Unit leaders should consider how teaching is weighted in formal evaluation processes in comparison to research and service for faculty of different ranks and appointments. If the unit values teaching but does not reward it in hiring, promotion, or tenure processes unit leaders and faculty should work together to revise those evaluation processes. Unit leaders can also use teaching awards and public recognition to incentivize and reward teaching improvement. Even informally spotlighting faculty's assessment and improvement work during meetings or in electronic communications can demonstrate that teaching is valued. In addition to highlighting individual faculty's work, leaders should also communicate unit-level assessment of teaching trends to faculty. Unit-level data can be used to discuss how results will inform decision-making and faculty development planning.

Leaders' support, particularly when it comes to revising faculty evaluation process to compensate faculty for teaching improvement, may facilitate the cultural change necessary for some units to enact their plans (Andrews et al., 2020; Eckel et al., 1999; Kezar, 2018; Weaver et al., 2020).

In addition to cultivating support within, units need continued direction and support from the university's administration. The document analysis of assessment of teaching plans revealed several patterns potentially related to how well units understood expectations for assessment of teaching. One unit that submitted its plan before support resources were available, for example, had the lowest overall rating of all the plans, likely due to a lack of understanding of what to include in the plan. Units have also contacted me to request information about next steps in this process. Institutional leaders should work with ACIE, unit leaders, the faculty senate, and the Teaching Center to determine how assessment of teaching work will be tracked, reported, and used in the future and communicate this information to faculty in a clear and timely manner. Ongoing tracking and reporting also creates opportunities for institutional leaders to enhance the visibility of assessment of teaching work by calling attention to units' teaching assessment and improvement project successes (Weaver et al., 2020).

Units that consulted with the Assessment of Teaching Initiative tended to have more comprehensive assessment of teaching plans. Focus groups reported finding that the Teaching Center was a helpful resource in assessment of teaching revision planning. In addition to the existing, institution-wide resources that the Initiative has created, the Assessment of Teaching Initiative should continue to advertise and provide individualized support in the form of curated resources, consultations, feedback on assessment processes and tools, and unit-specific faculty development events to units. Research from the TEval project suggests that it may be beneficial for the Assessment of Teaching Initiative to assist with and facilitate unit assessment of teaching

meetings to provide units with embedded support and help sustain change momentum (Andrews et al., 2020; Weaver et al., 2020). The Teaching Center will need to strategize how to communicate options for resources and support to academic units that may be less aware of the services. Beyond providing units with valuable expertise, collaborating with the Assessment of Teaching Initiative may save faculty time and effort and help units plan how to connect faculty development to assessment of teaching. The Teaching Center can also publicize units' successes, which may increase the visibility and valuation of assessment of teaching work.

As units begin to implement their assessment of teaching plans, they should collect outcomes data to inform iterative revision to assessment processes. Focus group participants identified potential outcomes. Participants reported that assessment of teaching processes must be feasible and sustainable to be successful. They also stated that they believed planned revisions would create a more equitable teaching evaluation process for faculty and generate higher quality data. Units should collect faculty feedback and self-reported data to determine whether the changes their units made constitute improvements and to assess feasibility and sustainability. Units will also need to examine unit-level assessment of teaching data to track changes in teaching effectiveness over time and create recommendations for faculty development and teaching improvement.

In a decentralized university, enacting widespread continuous improvement changes creates the potential for siloing and for the focus to shift from improvement to compliance. To foster cross-unit collaboration, institutional leaders and the Assessment of Teaching Initiative should leverage existing events like Pitt's annual Assessment and Teaching Conference to create opportunities for units to share information, resources, progress, and lessons learned. Sharing information facilitates organizational learning necessary to advance assessment of teaching

improvement at an institutional level. The Assessment of Teaching Initiative should also expand resources and services to encourage ongoing formative assessment of teaching that fosters continuous teaching improvement.

5.2.2 At Other Institutions

The TEval project (Andrews et al., 2020; Finkelstein et al., 2020; Weaver et al., 2020) created model for revising teaching evaluation using the objectives of the Teaching Quality Framework (TQF), which involved developing a broad, institution-wide definition of teaching effectiveness that is then contextualized at the department-level. This model has the benefit of establishing a standard set of criteria and process for developing assessment of teaching. Universities that adopt the TQF follow a process of (Andrews et al., 2020):

- adopting an institutional-level definition of teaching effectiveness
- continuously cultivating interest in change within units
- building change teams
- facilitating ongoing team meetings to develop shared values and a vision of teaching effectiveness
- supporting teams as they create assessment processes and tools, pilot them, and revise them

Starting with a shared understanding of the dimensions of good teaching and a procedure for managing change creates consistency and gives change agents a higher degree of control over the process but sacrifices some flexibility and academic unit autonomy. For example, Finkelstein et al. (2020) noted that one of the challenges departments frequently experienced was confusion

about how to apply the Teaching Quality Framework rubric to assess teaching. This may be because the department did not develop the framework and rubric.

Pitt's approach gave academic units the ability to determine their own approaches and definitions for teaching effectiveness, it also led units to create assessment of teaching plans that varied significantly. Pitt's process involved:

- a multi-disciplinary committee of teaching experts developing broad recommendations for units to create assessment of teaching plans;
- the Office of the Provost establishing a timeline for plan creation;
- academic units developing definitions of teaching effectiveness and processes and tools for assessing teaching with support from the Assessment of Teaching Initiative;
- units submitting and receiving feedback from the Assessment of Teaching Initiative on plans;
- units piloting plans with support from the Assessment of Teaching Initiative as requested.

None of the 10 plans analyzed met the standard for comprehensiveness. Pitt's process and the Teaching Center's service model also constrained me to providing units with support by request rather than facilitating change more proactively. Instead of following a standard change process that allowed me to collaborate closely with units, I met with units that reached out, conducted needs assessments, and developed sets of resources and strategies. Ultimately, the units that chose to work with the Assessment of Teaching Initiative received customized assistance, but most units did not request support. Higher education institutions that embark on revising assessment of teaching will need to determine the balance of flexibility and unit autonomy with consistency and quality control that is appropriate for their institution. For large or decentralized higher education institutions or institutions with units with varying levels of experience for or interest in conducting

multiple measures of teaching, it may be beneficial for more decision-making to occur at the unit level.

Institutions preparing to begin revising their assessment of teaching should consider how various stakeholders can support this type of multi-level change. TEval and Pitt can serve as two examples of instigating institutional revision of assessment of teaching, however, other universities need to determine how best to plan and manage this type of change with attention to their own specific institutional and unit contexts. Members of the institution's administration should apply the following questions at the institutional level. Unit leaders and faculty should consider these questions at the unit level.

- What is the university's or unit's readiness for change?
 - What is the structure of the university or unit? How are decisions made? Who makes them?
 - What types of summative and formative assessment of teaching are already occurring at the institution or within the unit?
 - How well does revising assessment of teaching align with institutional or unit culture, goals, and values?
 - What are leaders' and faculty members' attitudes toward this change? What are their concerns? How might you address those concerns?
 - What policies, processes, structures, and tools would need to be revised for this change to take place?
 - What supports, centralized or within the unit, exist to support this change?
 - What are regional and specialized accreditation standards relevant to this change at your institution?

- How will various stakeholders cultivate and sustain support for this change?
- Which stakeholders should designate the process and timeline?
- Should change occur simultaneously in units or be rolled out gradually?
- Should the institution adopt one definition of teaching effectiveness that is contextualized at the unit-level or should units develop their own definitions? Who should participate in articulating a definition?
- Should there be a standard change process or should units develop their own? Who should participate in designing a process?
- Who supports provides units with expertise, guidance, training, and feedback throughout this process?
- How will this process be tracked? Who tracks it?
- Who should report on unit progress? How?
- How will information about the process and timeline for revising assessment of teaching be communicated to stakeholders?
- How can various stakeholders incentivize and reward engaging in this process?
- Who, within units, should manage and facilitate change planning and implementation? Leaders or faculty?
- Beyond improving teaching effectiveness, what should be the outcomes of revising assessment of teaching?

Teaching center or other faculty development staff should consider their department's mission, goals, and capacity to support this type of change. Many teaching centers do not support unit-level assessment of teaching and will need to determine whether expanding their scope of services to encompass this type of service is desirable and feasible. Teaching centers can curate

resources, support plan development, and provide consultations and training. Teaching centers can also potentially play a role in tracking, analyzing, and providing feedback on assessment of teaching plans, tools, and results. Perhaps most importantly, teaching centers are well-positioned to advocate for the use of formative assessment of teaching to foster faculty reflection, discussion, and teaching improvement and to help units use data to plan training and development. In this way, teaching centers can help maintain the focus on improvement rather than measurement for measurement's sake. The type and level of support should fit the function and capacity of the institution's teaching center, but the findings of this dissertation indicate that teaching centers are a potentially untapped resource for helping units revise assessment of teaching.

5.3 Limitations and Suggestions for Future Research

This research was not without limitations. First, as they were designed to be concise first drafts, assessment of teaching plans may not have captured all relevant assessment of teaching work or changes that a unit plans to or ultimately implements. Additional analysis of unit's policies and reports on assessment of teaching could supplement plan analysis to create a more complete picture of assessment of teaching comprehensiveness in units. The small number of focus groups was also a limitation as those selected likely did not capture the perspectives of faculty in units somewhere between just beginning to and relatively experienced at conducting multiple measures of teaching effectiveness. Focus groups participants were not representative of all faculty within their units. For example, only one focus group participant was a non tenure-stream faculty member. Additionally, because some had worked with me on assessment of teaching, may have responded in socially desirable ways to focus group questions. Collecting faculty perception data from a

broader, more representative group including part-time and non tenure stream faculty using a tool like surveys could provide new perspectives that did not emerge during the focus groups.

The nature of this research was subjective and interpretive. This could be advantageous and could create limitations. My work in the Assessment of Teaching Initiative provided me with experience, expertise, and insights that informed my interpretation of this data. When analyzing focus group data, for example, my experiences working with those two units provided important information about the context and timeline for revising assessment of teaching and the units' cultures. My involvement in this work may have also created subconscious biases that affected the way I coded and interpreted focus group data. Using a second coder to analyze data could provide additional, valuable perspectives and improve the trustworthiness of the data.

Lastly, given the timing of this dissertation of practice, the COVID-19 pandemic almost certainly affected assessment of teaching revision planning and potentially, focus group participants' responses. Academic units and faculty participated in this study while balancing a myriad of competing priorities, including managing an abrupt transition to fully online instruction. Repeating these measures later, when faculty members have more time to devote to assessment of teaching work or conducting longitudinal research on how plans evolve over time would likely yield more comprehensive plans and differences in focus group question responses.

Although assessment of teaching research is growing thanks in part to research on issues with teaching survey data and increasing faculty support for using other methods of assessment, this dissertation of practice represents one of the few studies that analyzed university units' approaches to and faculty perspectives on revising assessment of teaching. Research is needed to identify specific strengths and weaknesses of various approaches and effective strategies for addressing barriers to revising assessment of teaching. Additional studies should be conducted to

examine the outcomes of assessment of teaching revision to determine whether plan comprehensiveness affects outcomes and which types of strategies and methods best support improved teaching and learning and equitable teaching evaluation.

5.4 Conclusion

The results of this study suggest that unit culture and resources can facilitate or impede revision work. Specific drivers and barriers and how they affect revising assessment of teaching vary in different contexts, meaning that change must be planned and managed with attention to institutional and unit contexts. Institutions should identify resources to build capacity and strategies to support cultural change necessary to support revising assessment of teaching (Eckel et al., 1999; Finkelstein et al., 2020; Kezar, 2018; Weaver et al., 2020). This involves determining the extent to which the institution, its leaders, and its faculty value, support, and tangibly reward teaching work and improvement (Arreola, 2007; Bradforth, et al., 2015; Finkelstein et al., 2020; Weaver et al., 2020). Although the results of this dissertation suggest that many faculty will be amenable to changes, institutions should still anticipate and address faculty concerns and points of resistance. Leaders should consider how to structure assessment of teaching work to maximize faculty involvement while minimizing the burden to faculty, which may include creating or utilizing teams of champions (Eckel et al., 1999; Finkelstein et al., 2020; Kezar, 2018; Reinholz et al., 2017; Weaver et al., 2020). Institutions may be able to leverage faculty members' desire to stop overusing teaching surveys to cultivate broad faculty support for revision of assessment of teaching (Finkelstein et al., 2020; Weaver et al., 2020). Taking the time to develop faculty and support is crucial for the success of this type of change.

For most institutions, comprehensively revamping teaching assessment and improvement requires a transformational change that necessitates more than revising policies, processes, and tools. Teaching survey overuse is a deceptively straightforward and widely recognized problem in higher education, which has led to many organizations and experts calling for universities to address this issue (ASA, 2019; AAUP, 1975; Arreola, 2007, Berk, 2006; Burdsal & Harrison, 2008; Dennin et al., 2017; Esarey & Valdes, 2020; Miller & Follmer Greenhoot, 2008; Taylor, 2021; TEval, n.d.; Uttl et al., 2017; Weaver et al. 2020). But the ways in which universities assess and improve teaching are indicative of the underlying cultures of our institutions. Meaningful change will require examining and deconstructing the norms, values, assumptions, and attitudes that contributed to this problem. Focusing on solutions prematurely, without examining the system in which the change will occur and the levels of change involved may lead to compliance-focused or unsustainable assessment of teaching. Assessing teaching more effectively and using data to inform continuous improvement of teaching is possible and necessary but should be undertaken strategically as a complex, continuous organizational change process.

6.0 Reflection

In 2017, Montenegro and Jankowski published what would become a seminal work in assessment of student learning outcomes literature, the National Institute of Learning Outcomes Assessment (NILOA) occasional paper, *Equity and Assessment*. Equity should not have been a new topic in outcomes assessment. Teachers have been arguing for and practicing equitable teaching strategies for decades. Yet, it was not until four years ago that NILOA scholars proposed that assessment should identify and help address disparate outcomes. The authors argued that universities have an obligation give all students the support that they need to achieve equitable outcomes. Montenegro and Jankowski (2017) exhorted assessment professionals to scrutinize methods, data, and our own positionality to consider who has and has not benefitted from teaching improvement work. Examining how assessment is conducted and used can reveal some of the ways in which inequity is perpetuated in universities.

One of the most glaring examples in teaching assessment is the overuse of teaching surveys in faculty evaluations. As the program manager of the Assessment of Teaching Initiative, I have worked with minoritized faculty who had received biased student feedback and have grown to dread reading or discussing their teaching survey results. As a part-time instructor, I value student feedback but have feared that a one-time student survey that did not accurately or adequately portray my teaching could be used as the only evidence of the quality of my teaching. Research on assessment of teaching clearly indicated that this was not an isolated or new issue. The debate around whether teaching surveys are problematic has largely been settled; more recent studies have shifted focus toward investigating teaching surveys' specific shortcomings. Assessment of teaching researchers, faculty, and organizations that work to improve teaching like the AAU, agree

that using multiple assessment methods and triangulating results is a more effective, and often more equitable, way to assess teaching (AAUP, 1975; Arreola, 2007; Benton & Cashin, 2014; Berk, 2006; Vasey & Carroll, 2016). I wondered why, if there was widespread acknowledgement of the problem and known solutions, higher education institutions continued to over rely on teaching surveys. What stops us from adopting methods that better serve our faculty and students? When the provost and ACIE released the timeline and recommendations for academic units to develop assessment of teaching plans, it created an opportunity to begin to address this question at Pitt.

Examining an institutional change process was made more challenging by timing. I had to simultaneously build the Assessment of Teaching Initiative while assisting and encouraging teaching assessment revision. Due to delays created by the COVID-19 pandemic, academic units were ultimately given one semester, spring 2021, to create assessment of teaching plans. The unit leaders and faculty who developed these plans were, at the time, grappling with remote instruction, the increased emotional labor of supporting students through trauma, and their own stressful life and family situations. Managing change through crisis required responding to constantly shifting deadlines, priorities, and needs. Helping faculty during this time reinforced the necessity of prioritizing equity and care when supporting institutional change initiatives. Determining faculty members' needs and addressing their concerns and challenges had to precede planning and implementing assessment of teaching improvements.

Mays Imad has written extensively about trauma and care-informed pedagogy. In spring 2021, she published an article describing trauma-informed educational development. Reading it, I recognized many of the strategies I had used to facilitate assessment of teaching projects, like reducing unnecessary administrative work, fostering clear communication, helping build

communities of faculty who are passionate about teaching within units, and working with units to create assessment of teaching plans designed to be equitable for faculty and to encourage inclusive teaching. Facilitating change with care requires giving stakeholders agency and voice in processes and policies that affect them. This dissertation of practice taught me that middle-out approaches led by diverse, representative groups with a shared goal can be one of the most effective ways to orchestrate change in a university. Many of the problems that emerged were due to faculty anxiety or resistance to creating new, potentially more time-consuming assessment methods that would be used in decision-making processes that still lacked transparency or fairness. In units that took middle-out approaches, faculty were more engaged throughout plan development and often voiced and addressed concerns proactively. Although middle-out approaches are less linear and generally take more time than top-down decision-making, leveraging champions to build support and incorporating stakeholder feedback and expertise produces changes that are meaningful and more likely to be sustained. Close collaboration with diverse groups of instructors working to transform the way that we assess teaching reminded me that care--for faculty's teaching experiences and students' learning experiences--is a central but often overlooked component of teaching improvement.

Practicing inquiry and studying organizational change, particularly during a time characterized by so much transition that the phrase “the new normal” entered our collective lexicon, prompted me to think critically about the purpose of assessment and why practice often misaligns with purpose. If assessment is meant to facilitate equitable improvement, using new measures and metrics for evaluation matters less than providing feedback, resources, and structures to support development. I recently read the work of a professor who replaced instructor-assigned grades with reflective student self-assessments. He argued that “grades are anathema to the

presumption of the humanity of students, support for their basic needs, and engaging them as full participants in their own education” (Stommel, 2021, n.p.). To what extent does faculty evaluation support faculty needs or engage their humanity-as full participants in their own development? Academic units have created new processes and tools, but without interrogating the culture that created and sustained ineffective, inequitable assessment of teaching for decades, changes might not actually change much. In supporting assessment and improvement work with care, I hope to explore how to measure and evaluate less and create more opportunities for reflection and growth.

Appendix A Document Analysis Rubric

Plans needed to meet all criteria at the level of “developing” or higher to be deemed comprehensive.

| Performance criterion | Levels of performance | | |
|---|--|--|---|
| | Information missing or criterion not met | Developing | Accomplished |
| Context-specific definition of teaching effectiveness | | Plan contains a definition that is vague or lists characteristics of teaching effectiveness that cannot be measured and/or characteristics of faculty work that are unrelated to teaching. | The plan contains a context-specific definition that includes specific, measurable characteristics of teaching effectiveness. |
| Faculty involvement | | At least some faculty had opportunities to be involved in the creation of or provide input on the plan, but not in a substantial or sustained way. | Many faculty or a group of representative faculty (e.g. committee with representatives from each department) in the unit or were substantially involved throughout the plan creation process. |
| Multiple, varied sources of evidence for summative assessment | | <p>The plan lists and describes 1-2 methods of summative assessment of teaching that involve the collection of evidence from sources other than students (e.g. peer review, review of teaching or student learning artifacts, and/or instructor self-assessment).</p> <p>Methods and tools may not be fully developed at this point.</p> | <p>The plan lists and describes the several methods of summative assessment of teaching that allow for the collection of evidence from the following, varied sources: peer review, teaching surveys or another form of student feedback, review of teaching or student learning artifacts, instructor self-assessment.</p> <p>Although they may not be fully developed at this point, the plan addresses the next steps or timeline the unit will take to create methods and tools.</p> |
| Addresses formative assessment and connects to faculty development | | <p>The plan lists and describes at least one method of formative assessment of teaching that the unit will encourage or make available to instructors for the purpose of improving rather than evaluating teaching.</p> <p>Formative assessment methods are not explicitly tied to faculty development in the plan.</p> | <p>The plan lists and describes multiple methods of formative assessment of teaching that the unit will encourage or make available to instructors for the purposes of improving rather than evaluating teaching. The plan describes how instructors can request formative feedback from peers or mentors in the unit.</p> <p>Formative assessments are explicitly tied to faculty development in the plan.</p> |

| | | | |
|--|--|---|--|
| | | | The plan states that the unit will provide instructors with resources and training as requested or needed. |
| Tools align with the unit's definition of teaching effectiveness | | The plan includes assessment tools that include criteria that at least partially align with the unit's definition of teaching effectiveness. The plan does not indicate that the unit intends to revise the tools to better align with the definition. | The plan includes assessment tools with criteria that align with the definition of teaching effectiveness OR the plan indicates that the unit intends to revise the tools to align with the definition. |
| Reviewers and reviewer preparation | | The plan identifies who will conduct each type of assessment (e.g., committee, peer, mentor, department chair), but does not indicate that reviewers will receive training. | The plan identifies who will conduct each type of assessment (e.g., committee, peer, mentor, department chair) and indicates that reviewers will receive training to ensure the consistency and equity of assessments. |
| Flexibility to account for differences in rank/appointment, courses, methods of delivery, and/or faculty teaching style | | The plan indicates that the unit differentiates or plans to differentiate assessment of teaching to some degree based on instructor rank/appointment, course type, and/or method of delivery. Differentiation is minor and/or does not provide instructors with a significant degree of flexibility when it comes to selecting sources of evidence. | The plan indicates that instructors have a significant degree of flexibility when it comes to selecting sources of evidence to demonstrate teaching effectiveness. The unit may provide instructors with a list of acceptable sources of evidence from which they may choose. |
| Use of results for continuous improvement of teaching | | The plan describes how the unit will use assessment of teaching data to inform continuous improvement of teaching at the instructor-level and mentions without describing thoroughly that results will also be used for unit-level continuous improvement. | The plan thoroughly describes how the unit will use assessment of teaching to inform continuous improvement of teaching at the instructor- and unit-levels. This description addresses who in the unit will have access to and review data, how results are communicated and shared, and the process for translating results into recommendations for improvement. |

Appendix B Focus Group Interview Protocol and Questions

Appendix B.1 Protocol (adapted from Dos Santos Marques et al., 2020)

1. Prior to the virtual focus groups:
 - a. Choose Zoom, a university-supported, secure virtual platform to conduct virtual focus groups.
 - b. Identify eligible, willing participants from the groups of faculty members who contributed to units' assessment of teaching plans.
 - c. Email participants details about the focus group purpose and design and a link to a Doodle poll that they could use to indicate their availability.
 - d. Schedule the focus group session using password protection and required institutional authentication in Zoom.
 - e. Email reminders and instructions for accessing the meeting to participants.
2. Day of the virtual focus group session:
 - a. Record session attendance.
 - b. Welcome participants and perform introductions.
 - c. Describe the purpose of the study.
 - d. Describe recording and confidentiality.
 - e. Obtain verbal consent.
 - f. Explain discussion guidelines.
 - g. Invite questions.
 - h. Turn on Zoom recording.

3. Follow up:
 - a. Email participants to thank them for their participation.

Appendix B.2 Questions

1. What were your school's goals for revising assessment of teaching processes?
2. Describe your school's process of revising assessment of teaching and creating an assessment of teaching plan.
 - a. What were the steps in the process?
 - b. What was your timeline?
 - c. Who was involved? How and to what extent were they involved?
3. Consider your school's approach and faculty involvement in this process. Was your process more top-down or bottom-up? Why?
4. What institutional factors (structures, processes, policies, aspects of institutional culture, resources) have affected the creation and/or revision of assessment of teaching processes? How?
5. If you could change or eliminate one of the institutional factors you identified to facilitate creating or implementing assessment of teaching, what would you change? Why?
6. Reflect on how your school assessed teaching prior to planning/instituting multi-method assessment of teaching (defined as an approach that involves triangulating data from multiple measures to both provide faculty with feedback for improvement

- and conduct evaluation of teaching). How do you think changing the way you measure teaching over the past several years will affect/has affected:
- a. equity of assessment of teaching?
 - b. completeness and quality of the teaching effectiveness data you gather?
 - c. the way you use data to improve teaching?
7. What, if anything else, would you like me (as someone who supports academic units in this process) to know?
8. What, if anything, would you share with institutional leaders to aid in continuous improvement of assessment of teaching at Pitt?

References

- AAUP Committee on Teaching, Research, and Publication (1975). *Statement on teaching evaluation*. American Association of University Professors.
<https://www.aaup.org/report/statement-teaching-evaluation>
- American Sociological Association. (2019, September). *Statement on student evaluations of teaching*. American Sociological Association.
https://www.asanet.org/sites/default/files/asa_statement_on_student_evaluations_of_teaching_sept52019.pdf
- Andrews, S.E., Keating, J., Corbo, J.C., Gammon, M., Reinholz, D.L., & Finkelstein, N. (2020). Transforming teaching evaluation in the disciplines: A model and case study of departmental change. In K. White, A. Beach, N. Finkelstein, C. Henderson, S. Simkins, L. Slakey, M. Stains, G. Weaver, & L. Whitehead (Eds.), *Transforming institutions: Accelerating change in higher education* (Ch. 13). Pressbooks,
<http://openbooks.library.umass.edu/ascenti2020/chapter/andrews-et-al/>
- Argyris, C. (1999). *On organizational learning* (2nd ed.). Blackwell Business.
- Arreola, R.A. (2007). *Developing a comprehensive faculty evaluation system: A guide to designing, building, and operating large-scale faculty evaluation systems*, (3rd ed.). Anker Publishing Company, Inc.
- Barlow, K.K. (2017, Mar. 16). Group wants change in how teaching is evaluated. *University Times*, 49(14). <https://www.utimes.pitt.edu/archives/?p=42833>
- Benton, S.L., & Cashin, W.E. (2014). Student ratings of teaching: A summary of research and literature. IDEA Paper #50. *IDEA*.
https://www.ideaedu.org/Portals/0/Uploads/Documents/IDEA%20Papers/IDEA%20Papers/PaperIDEA_50.pdf
- Berk, R.A. (2006). *Thirteen strategies to measure college teaching: A consumer's guide to rating scale construction, assessment, and decision-making for faculty, administrators, and clinicians*. Stylus Publishing, LLC.
- Bowen, G. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27-40. <http://dx.doi.org/10.3316/QRJ0902027>
- Boyce, M.E. (2003). Organizational learning is essential to achieving and sustaining change in higher education. *Innovative Higher Education*, 28(2), 118-136.

- Bradforth, S. E., Miller, E. R., Dichtel, W. R., Leibovich, A. K., Feig, A. L., Martin, J. D., Bjorkman, K. S., Schultz, Z. D., & Smith, T. L. (2015). University learning: Improve undergraduate science education. *Nature*, 532(7560), 282–284. <https://doi.org/10.1038/523282a>
- Brownell, S.E., & Tanner, K.D. (2012). Barriers to faculty pedagogical change: Lack of training, time, incentives, and...tensions with professional identity?. *CBE Life Sciences Education*, 1(4), 339–346. <https://doi: 10.1187/cbe.12-09-0163>
- Burdsal, C.A., & Harrison, P.D. (2008). Further evidence supporting the validity of both a multidimensional profile and an overall evaluation of teaching effectiveness. *Assessment & Evaluation in Higher Education*, 33(5), 567-576. <https://doi: 10.1080/02602930701699049>
- Campbell, J.P., & Bozeman, W.C. (2007). The value of student ratings: Perceptions of students, teachers, and administrators. *Community College Journal of Research and Practice*, 1. <https://doi.org/10.1080/10668920600864137>
- Clark, B. R. (1983). *The higher education system: Academic organization in cross-national perspective*. University of California Press.
- Cudd, A. (2021, February 16). *Memo on assessment of teaching*. Office of the Provost, University of Pittsburgh. <https://www.provost.pitt.edu/announcements/provost-memos/memo-assessment-teaching-february-16-2021>
- Cummings, R., Philips, R., Tilbrook, R., & Lowe, K. (2005). Middle-out approaches to reform of university teaching and learning: Champions stringing between the “top-down” and “bottom-up” approaches. *International Review of Research in Open and Distance Learning*, 6(1). <https://files.eric.ed.gov/fulltext/EJ846611.pdf>
- Dennin, M., Feig, A., Finkelstein, F., Follmer Greenhoot, A., Hildreth, M., Leibovich, A.K., Martin, J.D., Miller, E.R., Moldwin, M.B., O’Dowd, D.K., Posey, L.A., Schultz, Z.D., & Smith, T.L. (2017). *Aligning practices to policies: Changing the culture to recognize and reward teaching at research universities*. Association of American Universities and the Research Corporation for Science Advancement. <https://www.aau.edu/sites/default/files/AAU-Files/STEM-Education-Initiative/Aligning-Practice-To-Policies-Digital.pdf>
- Dos Santos Marques, I. C., Theiss, L. M., Johnson, C. Y., McLin, E., Ruf, B. A., Vickers, S. M., Fouad, M. N., Scarinci, I. C., & Chu, D. I. (2021). Implementation of virtual focus groups for qualitative data collection in a global pandemic. *American Journal of Surgery*, 221(5), 918–922. <https://doi.org/10.1016/j.amjsurg.2020.10.009>

- Eckel, P., Green, M., Hill, B., & Mallon, W. (1999). *On change III: Taking charge of change: A primer for colleges and universities*. American Council on Education Project on Leadership and Institutional Transformation. <https://files.eric.ed.gov/fulltext/ED450621.pdf>
- Esarey, J. & Valdes, N. (2020). Unbiased, reliable, and valid student evaluations can still be unfair. *Assessment and Evaluation in Higher Education*, 45(8), 1106-1120. <https://doi.org/10.1080/02602938.2020.1724875>
- Fairweather, J.S. (2005). Beyond the rhetoric: Trends in the relative value of teaching and research in faculty salaries. *The Journal of Higher Education*, 76(4). <https://www.jstor.org/stable/3838846>
- Fan, Y., Shepherd, L. J., Slavich, E., Waters, D., Stone, M., Abel, R., & Johnston, E. L. (2019). Gender and cultural bias in student evaluations: Why representation matters. *PloS One*, 14(2). <https://doi:10.1371/journal.pone.0209749>
- Finkelstein, N., Andrews, S., Pawlak, A., Rezaei, D., Keating, J., Corbo, J., Reinholz, D., & Bernstein, D. (n.d.). *Teaching quality framework*. University of Colorado Teaching Quality Framework Initiative. <https://www.colorado.edu/teaching-quality-framework/>
- Finkelstein, N., Greenhoot, A. F., Weaver, G., & Austin, A. E. (2020). A department-level cultural change project: Transforming the evaluation of teaching. In K. White, A. Beach, N. Finkelstein, C. Henderson, S. Simkins, L. Slakey, M. Stains, G. Weaver, & L. Whitehead (Eds.), *Transforming Institutions: Accelerating Systemic Change in Higher Education* (Ch. 14). Pressbooks. <http://openbooks.library.umass.edu/ascenti2020/chapter/finkelstein-et-al/>
- Gillespie, K.J., & Robertson, D.L. (2010). *A guide to faculty development* (2nd ed.). Jossey-Bass.
- Golden, C., & Kirsch, L. (2020, May 18). *2019 recommendations from Provost's Advisory Council on Instructional Excellence and subsequent activities: An update for the Senate Educational Policies Committee* [PowerPoint slides]. Office of the Provost, University of Pittsburgh. <https://www.provost.pitt.edu/sites/default/files/Assessment%20of%20Teaching%20SEP%20Presentation%2018%20May%202020.pdf>
- Gormally, C., Evans, M., & Brickman, P. (2014). Feedback about teaching in higher ed: Neglected opportunities to promote change. *CBE-Life Sciences Education*, 13, 187-199. <https://doi:10.1187/CBE.13-12-0235>
- Goodrick, D. (2014). *Comparative case studies: Methodological briefs - Impact evaluation no. 9*. UNICEF Office of Research Methodological Briefs. http://www.dmeforpeace.org/wp-content/uploads/2017/06/Comparative_Case_Studies_ENG.pdf

- Halasek, K., Heckler, A., & Rhodes-Disalvo. (2020). Transforming the teaching of thousands: Promoting evidence-based practices at scale. In K. White, A. Beach, N. Finkelstein, C. Henderson, S. Simkins, L. Slakey, M. Stains, G. Weaver, & L. Whitehead (Eds.), *Transforming institutions: Accelerating change in higher education* (Ch. 20). Pressbooks, http://openbooks.library.umass.edu/ascenti2020/chapter/halasek_heckler_rhodes-disalvo/
- Harrell, D. (2018, Nov. 1). Are student evaluations the best measure of teaching effectiveness? *University Times*, 51(5). <https://www.utimes.pitt.edu/news/are-student-evaluations>
- Imad, M. (2021). Transcending adversity: Trauma-informed educational development. *To Improve the Academy: A Journal of Educational Development*, 39(3), 1-23.
- Kolb, S. (2012). Grounded theory and the constant comparison method: Valid research strategies for educators. *Journal of Emerging trends in Educational Research and Policy Studies*, 3(1), 83-86.
- Linse, A. R. (2017). Interpreting and using student ratings data: Guidance for faculty serving as administrators and on evaluation committees. *Studies in Educational Evaluation*, 54, 94-106. <https://doi:10.1016/j.stueduc.2016.12.004>
- McMurtrie, B. (2019, March 21). Many professors want to change their teaching but don't. One university found out why. *The Chronicle of Higher Education*. <https://www.chronicle.com/newsletter/teaching/2019-03-21>
- Mengel, F., Sauermann, J., & Zölitz, U. (2019). Gender bias in teaching evaluations. *Journal of the European Economic Association*, 17(2), 535-566. <https://doi:10.1093/jeea/jvx057>
- Miller, E., & Follmer Greenhoot, A. (2008, October). *New frameworks for documenting, evaluating, and improving faculty teaching* [PowerPoint slides]. Association of American Universities. <https://www.aau.edu/sites/default/files/Blind-Links/Teaching-Eval-AAU-Chairs-Oct2018.pdf>
- Mitchell, K. M. W., & Martin, J. (2018). Gender bias in student evaluations. *PS: Political Science & Politics*, 51(3), 648-652. <https://doi:10.1017/S104909651800001X>
- Montenegro, E., & Jankowski, N. A. (2017, January). *Equity and assessment: Moving towards culturally responsive assessment*. (Occasional Paper No. 29). University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA). <https://www.learningoutcomesassessment.org/wp-content/uploads/2019/02/OccasionalPaper29.pdf>

- Murphy, T., MacLaren, I., & Flynn, S. (2009). Toward a summative system for assessment of teaching quality in higher education. *International Journal of Teaching and Learning in Higher Education*, 20(2), 226-236. <https://files.eric.ed.gov/fulltext/EJ864340.pdf>
- Nasser, F., & Fresko, B. (2002) Faculty views of student evaluation of college teaching. *Assessment & Evaluation in Higher Education*, 27(2), 187-198. <https://doi: 10.1080/02602930220128751>
- Office of Institutional Research. (2021). *Fact book 2021*. University of Pittsburgh. <https://ir.pitt.edu/wp-content/uploads/2020/11/Fact-Book-2021.pdf>
- Office of the Provost. (2020). *Promotion and tenure reviews*. University of Pittsburgh Faculty Handbook. <https://www.facultyhandbook.pitt.edu/promotion-and-tenure-reviews>
- Peterson, D.A., Biederman, L. A., Andersen, D., Ditonto, T. M., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PloS One*, 14(5). <https://doi:10.1371/journal.pone.0216241>
- Plan for Pitt*. (2021). University of Pittsburgh. <https://www.planforpitt.pitt.edu/sites/default/files/plan-for-pitt.pdf>
- Reinholz, D. L., Corbo, J. C., Dancy, M. H., Deetz, S., & Finkelstein, N. (2015). Towards a model of systemic change in university STEM education. In G. C. Weaver, W. D. Burgess, A. L. Childress, and L. Slakey (Eds.), *Transforming institutions: 21st century undergraduate STEM education* (pp. 115–124). Purdue University Press
- Sacks, D., Payette, P., Ellis, D., & Volpe Horii, C. (2021). Leveraging centers for teaching and learning in accreditation review and quality assurance processes. *POD Speaks*, 4, 1-6.
- Saldana, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). Sage Publications.
- Smith, B. P., & Hawkins, B. (2011). Examining student evaluations of black college faculty: Does race matter? *The Journal of Negro Education*, 80(2), 149–162. <https://www.jstor.org/stable/41341117>
- Smith, B. P., & Johnson-Bailey, J. (2012). Student ratings of teaching effectiveness: Implications for non-white women in the academy. *Negro Educational Review*, 62-63(1-4), 115.
- Stommel, J. (2021, June 2). *Grades are dehumanizing; Ungrading is no simple solution*. Jesse Stommel. <https://www.jessestommel.com/grades-are-dehumanizing-ungrading-is-no-simple-solution/>
- Storage, D., Horne, Z., Cimpian, A., & Leslie, S. (2016). The frequency of "brilliant" and "genius" in teaching evaluations predicts the representation of women and African Americans across fields. *PloS One*, 11(3). [https:// doi:10.1371/journal.pone.0150194](https://doi:10.1371/journal.pone.0150194)

- Storberg-Walker, J. & Torraco, R. (2004, March 3-7). *Change and higher education: A multidisciplinary approach* [Paper presentation]. Academy of Human Resource Development International Conference (AHRD), Austin, TX.
<https://eric.ed.gov/?id=ED492430>
- Taylor, L.U. (Ed.). (2021). *Implications of race and racism in student evaluations of teaching: The hate u give*. Lexington Books.
- Transforming higher education – Multidimensional evaluation of teaching (TEval). (n.d.)
Resources. <https://teval.net/resources.html>
- University Center for Teaching and Learning. (2021). *Mission, vision, values*. University of Pittsburgh. <https://teaching.pitt.edu/mission-vision-values/>
- Uttl, B., White, C.A., & Wong Gonzalez, D. (2017). Meta-analysis of faculty’s teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54. 22-42.
<https://doi.org/10.1016/j.stueduc.2016.08.007>
- Valdez, A., Takahashi, S., Krausen, K., Bowman, A., & Gurrola, E. (2020). *Getting better at getting more equitable: Opportunities and barriers for using continuous improvement to advance educational equity*. WestEd.
<https://www.carnegiefoundation.org/wpcontent/uploads/2020/10/Getting-Better-at-Getting-More-Equitable-Opportunities-and-Barriers-for-Using-Continuous-Improvement-to-Advance-Educational-Equity.pdf>
- Vasey, C., & Carroll, L. (2016). How do we evaluate teaching? Findings from a survey of faculty members. *Academe*.
<https://www.aaup.org/article/how-do-we-evaluate-teaching#.YXjHmZ7MI2w>
- Weaver, G.C., Austin, A.E., Follmer Greenhoot, A., & Finkelstein, N.D. (2020). Establishing a better approach for evaluating teaching: The TEval project. *Change: The Magazine of Higher Learning*, 52(3), 25-31. <https://doi.org/10.1080/00091383.2020.1745575>
- Weick, K.E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21(1), 1-19.
- Zartman, I.W. (2005). Comparative case studies. *International Negotiation*, 10(1).
<https://doi.org/10.1163/1571806054741137>