TAKING A CLOSER LOOK AT THE TEACHER-CENTERED SYSTEMIC REFORM (TCSR) MODEL: IN-DEPTH EXPLORATIONS OF FACTORS INFLUENCING STEM INSTRUCTORS' INSTRUCTIONAL PRACTICES.

Lu Shi
Shaanxi, China
Master of Engineering, Xi'an Jiaotong University, China, 2017

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Chemistry

University of Virginia
May 2022

Committee members:
Dr. Marilyne Stains (Advisor)
Dr. David S. Cafiso
Dr. Cassandra L. Fraser
Dr. Rebecca R. Pompano
Dr. Lindsay Wheeler

Abstract

In the last decades, there have been numerous national calls for transforming instructional practices in Science, Technology, Engineering, and Mathematics (STEM) courses to enhance the retention of students interested in pursuing STEM majors and the level of science literacy of non-STEM majors. Despite these calls and associated efforts to propagate the usage of research-based instructional innovations, the majority of STEM faculty members are still using lecturing as the primary way of teaching STEM courses. The discipline-based education research community has uncovered several barriers and drivers to the adoption of these instructional innovations. However, more research is needed to understand the lack of propagation. In this dissertation, we leveraged the Teacher-centered Systemic Reform (TCSR) model to identify and explore different factors that could serve as barriers or drivers of instructional innovation.

The first project aimed at characterizing departmental climate around teaching, a contextual factor often posited as a barrier to instructional innovation by instructors, and its relationship to instructors' uptake of learner-centered instructional strategies. Results indicate that some elements that are essential to define departmental collective climate around teaching are lacking (e.g., policies, practices, expectations). Moreover, we found that psychological collective climate was not related to instructors' uptake of learner-centered instructional practice.

The second project discussed the evaluation of teaching practices, which is often proposed in the literature as a driver of innovation. In this study, we explored the complementarity of two instruments, the Classroom Observation Protocol for Undergraduate STEM (COPUS) and the Learner-centered Teaching Rubrics (LCTR), in measuring pedagogical practices. Analysis of the data indicated a partial misalignment between the two instruments (COPUS and LCTR) in measuring teaching practices and pointed out the importance to consider the purpose of evaluation when selecting the evaluation instrument.

Lastly, the third project focused on exploring the relationships between instructors' thinking about teaching and learning as well as their practices within the TCSR model. In particular, we explored general chemistry instructors' conceptions of assessment and their associated practices. The analysis of assessment practices showed that summative assessment tools were still the predominant means for instructors to evaluate students' learning in general chemistry. Instructors demonstrated a variety of rationales for assessing their students, however, most aligned with a summative perspective on assessment. We also found that instructors who recognized both summative and formative purposes for assessing students used more formative assessment. However, there were some inconsistencies between their conceptions about assessment and their practices which points to the complexity of measuring the connection between instructor thinking and instructional practices.

In summary, this dissertation explored the complex relationships among contextual factors, teacher thinking, and instructors' practices proposed by the TCSR model in teaching STEM courses. While significant advances have been made to understand barriers and drivers of instructional innovations, the results of these studies demonstrate the challenges and messiness in capturing factors influencing instructors' practices.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I'm glad that I made the decision to join Prof. Marilyne Stains' group and start my career doing chemistry education research. I would like to thank my advisor for being a wonderful mentor for supporting me to finish my degree, providing feedback on my progress, and giving advice for my life and work. Thanks, Marilyne, for providing the opportunity to work with you and be such a supportive and wonderful mentor!

My husband, Bo Zhang, thanks for your understanding and support when I decide to transfer to UVA. Thanks for being my husband, and as a team member of my life. My parents, thanks for your support and encouragement of any decisions that I made or anything that I want to do, even though you really want me to start a real job.

I would also like to thank current and previous lab members for their feedback and devoted time to help me revise and polish every presentation and paper. Especially, Annika Kraft and Dr. Dihua (Victoria) Xue, and Dr. Maia Popova for being wonderful academic sisters! Great thanks to the CER community who provided feedback and suggestions to guide my research.

Lastly, I thank all my friends, for the care and help no matter where I stay, especially, Ke Wei, Wei Wang, Dr. Jiaxi Li, Bowen Qin, Dr. Bowen Lou, and Dr. Zhe Li, thanks for being such amazing friends!

DEDICATION

To my Mom, Dad, and Husband,

Thanks for all your support and encouragement, that makes me who I am

***

谨以此献给妈妈,爸爸和爱人,

感谢你们一路来的支持和鼓励成就了今天的我

CHAPTER 1: Introduction

Retention Issues in STEM Fields

The President's Council of Advisors on Science and Technology points out that fewer than 40% of the students who intend to pursue a Science, Technology, Engineering, or Mathematics (STEM) major complete their degree within STEM fields (Olson & Riordan, 2012). Moreover, students from underrepresented groups are not well represented in these fields. Indeed, they account for approximately 70% of college students, but only 45% of the students in STEM (Olson & Riordan, 2012). There is thus a need to support members of these underrepresented groups in pursuing their degrees in STEM (Estrada, Burnett, Campbell et al., 2016).

Research indicated that most students tend to switch their STEM major to a non-STEM field or leave during their first two years of college (Almatrafi, Johri, Rangwala et al., 2017; Ohland, Sheppard, Lichtenstein et al., 2008; Olson & Riordan, 2012). Several reasons have been identified related to students' persistence. For instance, high school and freshman year GPAs are strong indicators of students' persistence in STEM fields (Mendez, Buskirk, Lohr et al., 2008). Self-efficacy and perceived ability in STEM were also found to be related to students' graduation rate in STEM (Cromley, Perez, & Kaplan, 2016). Another important aspect that has been identified by students is the teaching quality of the introductory STEM courses. Seymour and Hewitt stated in their book "*Talking about learning: Why undergraduates leave the sciences*" that 96% of the participating students mentioned ineffective teaching in their STEM courses as one of the reason for them to leave STEM majors (Seymour & Hewitt, 1997; Seymour, Hunter, Harper et al., 2019). Strenta (1994) also indicates similar findings from more than 5,000 students in science and engineering departments, indicating the quality of teaching is one of the reason that causes them to leave STEM fields.

Less research has been conducted to specifically focus on chemistry undergraduate students' retention. Adams interviewed 23 students who left or intended to leave the chemistry major from a large research-intensive university and indicated that more than 80% of the students leave the chemistry major because they had some issues with chemistry coursework which includes ineffective of teaching and assessment (Adams, 2016). As mentioned by the interviewed students, several students pointed to ineffective assessment as a reason for leaving chemistry major due to misalignment between teaching and assessment.

Recognizing the role of instructional practices in first year STEM courses in students' persistence in STEM fields, efforts have been conducted to develop and propagate new pedagogical innovations.

Slow Uptake of Pedagogical Innovations

Over the last decades, several new instructional strategies have been developed to better reflect our understanding of how students learn and address retention issues. These strategies were developed and supported by research to investigate the impact of pedagogical practices on students' learning which were defined as evidence-based instructional practices (EBIPs) (Rahman & Lewis, 2020). These strategies fall under the umbrella of active learning. A thorough review from psychology and discipline-based education research (DBER) teams from chemistry, biology, physics, engineering, astronomy, geography, and geoscience defined active learning as "a classroom situation in which the instructor and instructional activities explicitly afford students agency for their learning. In undergraduate STEM instruction, it involves increased levels of engagement with (a) direct experiences of phenomena, (b) scientific data providing evidence about phenomena, (c) scientific models that serve as representations of phenomena, and (d) domain-specific practices that guide the scientific interpretation of direct observations, analysis of data, and construction and application of model " (Lombardi, Shipley, Astronomy Team et al., 2021, p. 16).

Extensive research within and across STEM disciplines have demonstrated the positive impact of EBIPs/active learning on STEM student learning outcomes. The most cited mete-analysis comprised more than 200 studies published in education research journals in chemistry, physics, math, geology, engineering, biology, and computer science compared the effectiveness of using broadly defined active learning and lecturing on students' performance (Freeman, Eddy, McDonough et al., 2014). The results indicate that active learning increases examination performance by half of a standard deviation and lecturing increases failure rates by 55%. The findings echo one of the recent meta-analysis focused on understanding the effectiveness of utilizing EBIPs on student academic performance in chemistry (Rahman & Lewis, 2020). The study explored several EBIPs that have been researched during the year 2000 to 2017, including Cooperative Learning (Johnson, Johnson, & Smith, 1998), Collaborative Learning (Barkley, Cross, & Major, 2014), Problem-based Learning (PBL) (Dochy, Segers, Van den Bossche et al., 2003; Eberlein, Kampmeier, Minderhout et al., 2008), Process Oriented Guided Inquiry Learning (POGIL) (Moog & Spencer, 2008; Moog, Spencer, & Straumanis, 2006), Peer-led Team Learning (PLTL) (Wilson & Varma-Nelson, 2016), and Flipped Classrooms (Seery, 2015). The results from the meta-analysis which included 99 articles indicate that EBIPs can increase student academic performance compared to traditional lecturing-based classes with a range of an overall weight effect size from 0.29 to 0.62.

Although these meta-analyses have demonstrated the effectiveness of active learning in improving student academic performance, the majority of STEM faculty still use lecturing as their primary teaching methods. A study that collected classroom observations from more than 500 STEM faculty across 25 higher education institutions demonstrated that Didactic (i.e., on average, 80% of class time is spent lecturing) was the most prominent mode of instruction across the STEM disciplines (Stains, Harshman, Barker et al., 2018). A more recent large-scale survey study supports these findings (Benabentos, Hazari, Stanford et al., 2021). In that study, researchers developed a survey named as Change in Implementation of Pedagogical Practices (ChIPP) to assess teaching practices, institutional and departmental support for using student-centered teaching strategies, and faculty attitudes towards using innovative teaching strategies. They collected 1,456 surveys from faculty across 66 different institutions, with 48% from biology, 21% from chemistry, 18% from physics, and 13% from other STEM disciplines. Participants were asked to answer a question to report on the STEM courses that they teach most frequently at two points in time: the most recent occasion they taught it and the oldest occasion in the last five years. The listed teaching strategies consisted of 19 items, which included some instructor-centered strategies (e.g., lecturing, showing slides, or writing on the board) and some student-centered strategies (e.g., reflective activities, peer instruction, or some group problem-solving). The majority of the instructors (80%) reported using lecture, more than 70% of them showed slides, and around 50% written on the board in every class. The researcher also investigated the instructors' change of teaching practices over the five-year period. The results indicate that slightly over half of the faculty didn't change the frequency of using certain teaching strategies, while around 35% of the instructors who teach lower-division courses, and 28% of them who teach upper-division courses, have increased the relative frequency of use of student-centered strategies.

Research conducted in various Discipline-Based Education Research disciplines indicate that the lack of uptake of EBIPs is problematic in each STEM discipline. For example, the large observation-based study showed that chemistry courses had the highest proportion of lecturing compared to other STEM disciplines (Stains, et al., 2018). Indeed, 71% of the chemistry classroom observations were classified as didactic and only 12% were classified as student-centered. A more recent large-scale survey study echoes these findings (Raker, Dood, Srinivasan et al., 2021). Raker and co-authors surveyed more than 800 instructors from 4-year institutions within the United States to characterize their level use of active learning strategies, which included Peer-led team learning (PLTL), Problem-based learning (PBL), and Process-oriented guided-inquiry learning (POGIL) within the course that instructors have the most control. By following a rigorous sampling strategy and data analysis, the results indicate that around 17% of the participating instructors use PLTL and around 11% used POGIL and PBL. The lack of uptake points to issues with the effective dissemination of EBIPs. In the geosciences, a national survey was collected in 2004 from 2,207 geoscience instructors across the United States. It showed that 66% of the instructors rely solely on traditional lecturing in introductory courses and 56% use traditional lecturing nearly every class in courses for geoscience majors (Macdonald, Manduca, Mogk et al., 2005). Nearly half of the participating instructors incorporated some interactive teaching in the class including demonstrations, discussions, and in-class exercises. However, less than 10% of them reported that they use small group discussion, classroom debates/role-playing, or fieldwork which requires more student's interactions. A decade later, a team trained by the On the Cutting Edge sponsored Classroom Project re-evaluated the reformed teaching practices in geoscience (Manduca, Iverson, Luxenberg et al., 2017). They collected classroom observation data and instructor self-reported survey from around 200 instructors who teach geoscience across the United States. They analyzed the classroom observation data with the Reformed Teaching Observation Protocol (RTOP) (Sawada, Piburn, Judson et al., 2002) and classified instructors as Teacher-centered, Transitional, and Student-centered. The results of the analyses demonstrated that around 25% of participating instructors belonged to Student- centered, 30% were categorized as Teacher-centered, and 45% as Transitional. The triangulation of the classroom observation and survey data showed that only 10% of the instructors classified as Teacher-centered used small group activities during class and over 50% never used any EBIPs in their courses.

The physics education research community investigated the usage of EBIPs among physics instructors teaching introductory courses across different types of institutions in the United States. A web-based survey was distributed to 722 physics faculty in 2008 with the aim of characterizing introductory physics instructors' knowledge and level of use of 24 EBIPS that had been developed and disseminated by the physics education research community (Dancy & Henderson, 2010; Henderson & Dancy, 2009; Henderson, Dancy, & Niewiadomska-Bugaj, 2012). Analyses of the survey data indicated that 87% of the faculty reported that they were familiar with at least one of the 24 strategies, and around half of the instructors (49%) indicated familiarity of more than five of the strategies. Regarding the use of the EBIPs, 48% of the participants reported that they currently use at least one of the strategies, however, most of the instructors made some significant modifications when using EBIPs. Take the most used EBIP, Peer Instruction (PI), as an example: 41% of the respondents who use this strategy made significant modifications to it, and 6% of them who used PI were "not familiar enough with the developer's description." This lack of adherence to the original design of the EBIP may lead to diminished effectiveness.

In summary, this body of work indicates that STEM instructors have been slow to adopt EBIPs and active learning despite extensive evidence demonstrating their effectiveness in improving student learning outcomes. The question then is: why have STEM instructors not embraced these strategies? In the following section, we will describe research that aims to address this problem.

Challenges in Propagating Teaching Innovation

A large body of research has focused on characterizing the barriers and drivers to adopt EBIPSs among STEM instructors. Henderson (Henderson, Beach, & Finkelstein, 2011) did a thorough review back to 2011 included 194 literature from three types of researcher to understand what strategies are commonly used by change agents to promote instructional change in undergraduate STEM courses. The three group of researchers included: disciplinary-based STEM education researchers (SER), faculty development researchers (FDR), and higher education researchers (HER). Based on their research, four categories have been identified to capture change strategies: disseminating curriculum and pedagogy, focusing on developing reflective teachers, enacting policy, and developing shared vision. The results indicated that some overlaps have been identified among each category which point out the culture or norm of the organization need to be taken into consideration when conducting teaching innovation, and the faculty beliefs and conceptions about teaching and learning need to be addressed in more depth. They also summarized from the literature that the best practice for creating change requires long-term support for faculty, identifying faculty’s conceptions about teaching and learning, a reward system, as well as support within an institution to facilitate pedagogical innovation.

Sturtevant and Wheeler (2019) did a thorough review to identify the barriers to implement EBIPs, and categorized the barriers into four categories: teacher characteristics, student characteristics, technical/pedagogical issues, and institutional/departmental barriers. Under teacher characteristics which include lack of time for preparing EBIPs activities (Brownell & Tanner, 2012; Michael, 2007; Sunal, Hodges, Sunal et al., 2001; Turpen, Dancy, & Henderson, 2016), worries about content coverage (Dancy & Henderson, 2008; Turpen, et al., 2016), or loss of autonomy when teaching the course (Shadle, Marker, & Earl, 2017). Student characteristics include students’ resistance to participate in EBIPs activities (Dancy & Henderson, 2008; Walder, 2015), or students come to the class without enough preparation (Hastings & Breslow, 2015; Henderson & Dancy, 2007; Shadle, et al., 2017). Inadequate resources, such as, lack of TA for grading (Walder, 2015), inappropriate classroom layout (Dancy & Henderson, 2008), insufficient assessment methods for assessing students in active learning class (Shadle, et al., 2017) are identified as technical/pedagogical issues to implement EBIPs. The last barrier which emphasize on the institutional/departmental level include the lack of incentives (Brownell & Tanner, 2012; Chasteen, Perkins, Code et al., 2016; Elrod & Kezar, 2017), departmental norms did not favor EBIPs using (Dancy & Henderson, 2008), or insufficient faculty assessment to reward innovative teaching (Shadle, et al., 2017).

Regarding the drivers of implementing EBIPs, Shadle (2017) did a research at Boise State University from different STEM departments includes 169 instructors. Improving teaching and assessment has been identified as a crucial driver to implementing innovative teaching. It includes instructors expected to improve teaching ability and self-efficacy, as well as to have better approaches to assess their teaching and students' learning. Other top-rated drivers are 1) encouraging instructors to learn from the colleagues who had already successfully adapted innovative teaching; 2) encouraging collaboration and shared objectives to emphasize the importance of communication between instructors within the same department to develop a shared version of teaching; 3) aligns with existing resources which demonstrate the needs of resources (e.g., support from teaching learning center, technology demands) for implementing innovative teaching.

This body of work highlights the complexity in implementing instructional innovations among STEM faculty. While many drivers and barriers have already been investigated, more needs to be done to provide a complete picture. In the next section, we describe a theoretical framework that organizes barriers and drivers into categories. This categorization helps showcase relationships between different types of barriers/drivers and identifies areas needing of further investigation.

Teacher-Centered Systemic Reform (TCSR) model

In order to get a more holistic view of what factors will affect instructor' pedagogical practice, the Teacher-Centered Systemic Reform (TCSR) model has been proposed by Gess-Newsome and Woodbury as a framework to help understand factors influencing instructional practices (Gess-Newsome, Southerland, Johnston et al., 2003; Woodbury & Gess-Newsome, 2002). They conducted an exhaustive review of the literature which was used to guide educational reform at the K-12 level (Woodbury & Gess-Newsome, 2002). They then adapted the model for the higher education context (Gess-Newsome, et al., 2003). The framework proposes that teaching reform should consider instructors' beliefs about teaching and learning as well as their personal background within a situational teaching context. The TCSR model consists of four broad factors: Personal Factor, Teacher Thinking, Contextual Factor, and Instructor's Practice (Figure 1.1). The interconnections among the factors in this model explain the multi-faceted systemic nature of education and the need to consider these factors when a change/reform is to be implemented.

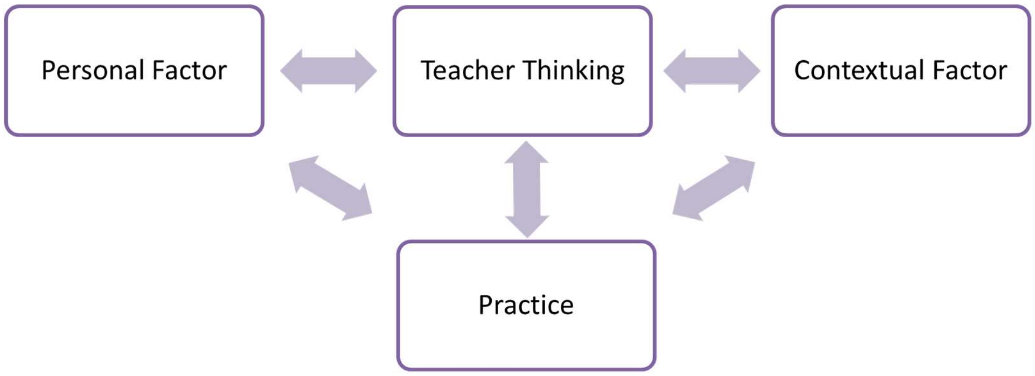


Figure 1.1 Overview of Teacher-Centered Systemic Reform (TCSR) model

Personal factor includes instructors' demographic information, current academic profile (e.g., position title, workload, job responsibilities), teaching experiences (e.g., TA, higher education, K-12, etc.), teachers' pedagogical training (e.g., attending teaching workshops, education degrees or certificates, etc.), teachers continued learning efforts (e.g., attending professional development workshops, consult with educational literatures, talk with colleague with teaching experts, etc.), and engagement in educational/bench research activities (e.g., engage in educational/bench research and publication, share educational/bench research results through seminar or presentation at a conference, etc.) (Gess-Newsome, et al., 2003; Woodbury & Gess-Newsome, 2002). Some literature in discipline-based education research supports the connection between Personal factor and Instructor's practice posited by the model. Lund and Stains (2015) studied the relationship between chemistry, biology, and physics instructors' exposure to EBIPs as students and their use of EBIPs. They surveyed and observed 99 instructors from different STEM departments at a Midwest research intensive institution to understand their awareness and adoption of EBIPs, attitudes and beliefs toward student-centered teaching, and their teaching practices. They found that instructors who had experienced EBIPs as students were more likely to use EBIPs in their own classroom.

The Teacher Thinking factor reflects the interdependent nature of knowledge and beliefs, denotes teachers' reflective, planning, and interactive thoughts, as well as concerns about teaching, teachers, learning, and learners. It consists of beliefs about teaching and learning, pedagogical content knowledge (PCK), self-efficacy, and dissatisfaction with current practices and students' learning. Existing literature illustrates the connections between instructors' thinking and instructional practices. Popova et al. interviewed 19 assistant chemistry professors about their beliefs about teaching and learning and used classroom observations as well as course artifacts to capture their instructional practices (Popova, Shi, Harshman et al., 2020). They found that instructors who hold student-centered beliefs are more likely to implement student-centered practices. However, there was no full alignment between beliefs and practices, and beliefs were more student-centered than their practices. Instructors' dissatisfaction with current teaching has also been identified as a driver for instructional innovation. Lemons and Andrews (2015) interviewed 17 biology instructors who attempt to incorporate case study in their teaching, and found that the instructors who were less likely to adopt active-learning strategies were satisfied with their lecturing strategies and believed that students can learn a lot from their lectures. On the contrary, instructors who plan to change their lecturing to more student-centered teaching strategies expressed dissatisfaction with their current pedagogical approach. Studies also point to the connection between the Personal factor and the Teacher Thinking factor. One study explored the impact of a short professional development program for new assistant professors in chemistry on their instructional practices, knowledge about instructional strategies, and self-efficacy - the latter two being part of the Teacher Thinking factor (Stains, Pilarz, & Chakraverty, 2015). They found that participation in the professional development program enhanced participants' knowledge of EBIPs and increased their self-efficacy in implementing these practices. Similar studies focused on impact of disciplinary professional development programs also indicate that pedagogical training of new faculty boost instructors' awareness and use of EBIPs (Derting, Ebert-May, Henkel et al., 2016; Henderson, 2008).

Contextual factor focuses on elements from the higher education system, including broader professional community, institution, department, and classroom/course context that could support or prevent teaching innovation. Course and classroom context has been widely studied to understand its connections to instructors' uptake of active learning (Apkarian, Henderson, Stains et al., 2021; Yik, Raker, Apkarian et al., 2022). STEM instructors attribute their lack of adoption to large class sizes or fixed classroom layout (Henderson & Dancy, 2007; Walczyk & Ramsey, 2003). One last aspect of Contextual factor that draws attention from the research community is the department level since the department serves as a key level to make pedagogical decisions (Corbo, Reinholz, Dancy et al., 2016; Reinholz & Apkarian, 2018). Serval research studies point to the departmental climate around teaching as a barrier to instructional change (Henderson & Dancy, 2007; Shadle, et al., 2017; Sturtevant & Wheeler, 2019). The ineffectiveness of teaching evaluation (e.g., student evaluation system) (Fan, Shepherd, Slavich et al., 2019; Hornstein, 2017) , as well as lack of policy to value teaching in the promotion process also serves as barriers to incorporate innovative teaching.

The goal of this dissertation was to explore further some factors from the TCSR model that have been described in the literature as influential in promoting instructional reforms.

Overview of the Dissertation

Leveraging prior research in drivers and barriers to adoption of instructional innovations and the TCSR model, this dissertation aims to provide insight into effective propagation strategies of EBIPs by exploring three factors within the TCSR model: 1) the relationship between departmental climate around teaching (Contextual factor) and adoption of EBIPs (Practice), 2) measures of teaching quality (Practice), and 3) instructors' conceptions of assessment (Teacher Thinking factor) and their assessment practices (Practice).

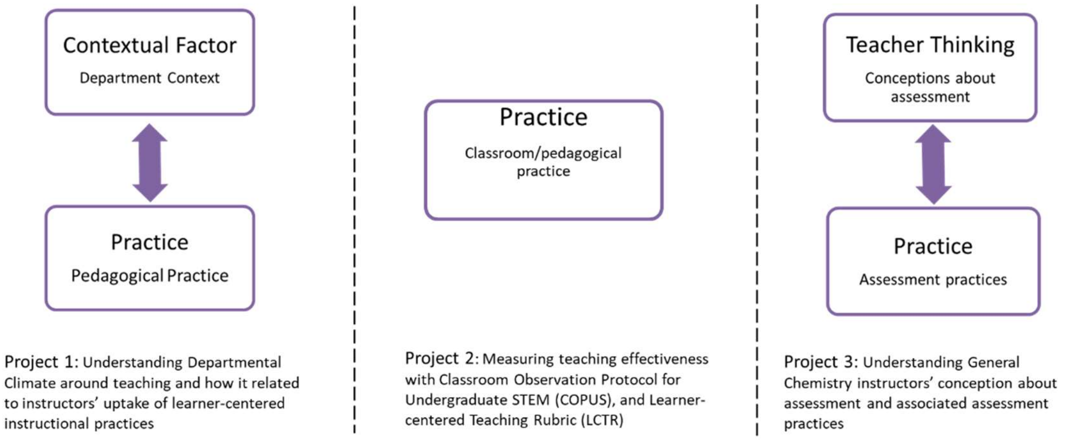

Contextual Factor
Department Context
Practice
Pedagogical Practice
Project 1: Understanding Departmental Climate around teaching and how it related to instructors' uptake of learner-centered instructional practices
Practice
Classroom/pedagogical practice
Project 2: Measuring teaching effectiveness with Classroom Observation Protocol for Undergraduate STEM (COPUS), and Learner-centered Teaching Rubric (LCTR)
Teacher Thinking
Conceptions about assessment
Practice
Assessment practices
Project 3: Understanding General Chemistry instructors' conception about assessment and associated assessment practices

Figure 1.2 Overview of the dissertation

Departmental climate around teaching has been identified in several studies surveying or interviewing STEM faculty as influential to instructional innovation (Shadle, et al., 2017; Emily Walter, Henderson, Beach et al., 2016). However, this body of work lacks an operationalization of departmental climate around teaching and few studies have empirically tested whether climate is actually related to implementation of innovative teaching practices. In Chapter two, we describe the development of a measure of departmental climate around teaching and the results of an exploration of the relationship between departmental climate and instructors' uptake of EBIPs.

The literature on instructional change also highlights that teaching evaluation could be a driver for instructors to adopt more learner-centered teaching strategies. New teaching evaluation frameworks prescribe the triangulation of evidence across various sources of data (Simonson, Earl, & Frary, 2021). While numerous measures of instructional qualities have been developed to provide this evidence, the landscape is complex and deciding on which set of measures to include can be confusing for instructors and administrators. In Chapter 3, we address this problem by exploring the complementarity of two different measures of instructional practices.

Lastly, chemistry-based education researchers have been advocating for the implementation of new assessment approaches that support the development of students' content knowledge and scientific practices (Bretz, 2014; Stowe & Cooper, 2019). In order to help the propagation of these new approaches, we first need to understand chemistry instructors' assessment practices and their rationales behind these practices. Chapter four addresses this gap in the literature by interviewing general chemistry faculty about their assessment practices and the rationales behind their practices.

The broad research questions that this dissertation addresses are:

1. To what extent does departmental climate around teaching relate to instructors' uptake of evidence-based instructional practices?
2. What is the complementary of two instruments designed to measure teaching practices?
3. What are general chemistry instructors' rationales behind their assessment practices?

References

Adams, G. (2016). Staying in chemistry: factors that predict undergraduate retention and recruitment at a top-ranked chemistry program and university.

Almatrafi, O.,Johri, A.,Rangwala, H., et al. (2017). *Retention and persistence among STEM students: A comparison of direct admit and transfer students across engineering and science.* Paper presented at the Proceedings of ASEE Annual Meeting.

Andrews, T. C.,Lemons, P. P. (2015). It's personal: Biology instructors prioritize personal evidence over empirical evidence in teaching decisions. *CBE—Life Sciences Education, 14*(1), ar7.

Apkarian, N.,Henderson, C.,Stains, M., et al. (2021). What really impacts the use of active learning in undergraduate STEM education? Results from a national survey of chemistry, mathematics, and physics instructors. *PloS one, 16*(2), e0247544.

Barkley, E. F.,Cross, K. P.,Major, C. H. (2014). *Collaborative learning techniques: A handbook for college faculty*: John Wiley & Sons.

Benabentos, R.,Hazari, Z.,Stanford, J. S., et al. (2021). Measuring the implementation of student-centered teaching strategies in lower-and upper-division STEM courses. *Journal of Geoscience Education, 69*(4), 342-356.

Bretz, S. L. (2014). Designing assessment tools to measure students' conceptual knowledge of chemistry. In *Tools of chemistry education research* (pp. 155-168): ACS Publications.

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, 11*(4), 339-346.

Chasteen, S. V.,Perkins, K. K.,Code, W. J., et al. (2016). The science education initiative: an experiment in scaling up educational improvements in a research university. *Transforming institutions: undergraduate STEM education for the 21st century*, 125-139.

Corbo,Reinholz,Dancy, et al. (2016). Framework for transforming departmental culture to support educational innovation. *Physical Review Physics Education Research, 12*(1), 010113.

Cromley, J. G.,Perez, T.,Kaplan, A. (2016). Undergraduate STEM achievement and retention: Cognitive, motivational, and institutional factors and solutions. *Policy Insights from the Behavioral and Brain Sciences, 3*(1), 4-11.

Dancy,Henderson. (2008). *Barriers and promises in STEM reform.* Paper presented at the National Academies of Science Promising Practices Workshop.

Dancy,Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American Journal of Physics, 78*(10), 1056-1063.

Derting, T. L.,Ebert-May, D.,Henkel, T. P., et al. (2016). Assessing faculty professional development in STEM higher education: Sustainability of outcomes. *Science Advances, 2*(3), e1501422.

Dochy, F.,Segers, M.,Van den Bossche, P., et al. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction, 13*(5), 533-568.

Eberlein, T.,Kampmeier, J.,Minderhout, V., et al. (2008). Pedagogies of engagement: A comparison of PBL, POGIL, and PLTL. *Biochem. Mol. Biol. Educ, 36*, 262-273.

Elrod, S.,Kezar, A. (2017). Increasing student success in STEM: Summary of a guide to systemic institutional change. *Change: The Magazine of Higher Learning, 49*(4), 26-34.

Estrada, M.,Burnett, M.,Campbell, A. G., et al. (2016). Improving underrepresented minority student persistence in STEM. *CBE—Life Sciences Education, 15*(3), es5.

Fan, Y.,Shepherd, L. J.,Slavich, E., et al. (2019). Gender and cultural bias in student evaluations: Why representation matters. *PloS one, 14*(2), e0209749.

Freeman, S.,Eddy, S. L.,McDonough, M., et al. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the national academy of sciences, 111*(23), 8410-8415.

Gess-Newsome, J.,Southerland, S. A.,Johnston, A., et al. (2003). Educational reform, personal practical theories, and dissatisfaction: The anatomy of change in college science teaching. *American Educational Research Journal, 40*(3), 731-767.

Hastings, D.,Breslow, L. (2015). Key elements to create and sustain educational innovation at a research-intensive university. *Transforming institutions: undergraduate STEM education for the 21st century*, 199-207.

Henderson,Dancy. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics-Physics Education Research, 3*(2), 020102.

Henderson, C. (2008). Promoting instructional change in new faculty: An evaluation of the physics and astronomy new faculty workshop. *American Journal of Physics, 76*(2), 179-187.

Henderson, C.,Beach, A.,Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching, 48*(8), 952-984.

Henderson, C.,Dancy, M. (2009). Impact of physics education research on the teaching of introductory quantitative physics in the United States. *Physical Review Special Topics-Physics Education Research, 5*(2), 020107.

Henderson, C.,Dancy, M.,Niewiadomska-Bugaj, M. (2012). Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovation-decision process? *Physical Review Special Topics-Physics Education Research, 8*(2), 020104.

Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education, 4*(1), 1304016.

Johnson, D. W.,Johnson, R. T.,Smith, K. A. (1998). Cooperative learning returns to college what evidence is there that it works? *Change: The Magazine of Higher Learning, 30*(4), 26-35.

Lombardi, D.,Shipley, T. F.,Astronomy Team, B. T., Chemistry Team, Engineering Team, Geography Team, Geoscience Team,, et al. (2021). The curious construct of active learning. *Psychological Science in the Public Interest, 22*(1), 16.

Lund, T. J.,Stains, M. (2015). The importance of context: an exploration of factors influencing the adoption of student-centered teaching among chemistry, biology, and physics faculty. *International Journal of STEM Education, 2*(1), 1-21.

Macdonald, R. H.,Manduca, C. A.,Mogk, D. W., et al. (2005). Teaching methods in undergraduate geoscience courses: Results of the 2004 On the Cutting Edge survey of US faculty. *Journal of Geoscience Education, 53*(3), 237-252.

Manduca, C. A.,Iverson, E. R.,Luxenberg, M., et al. (2017). Improving undergraduate STEM education: The efficacy of discipline-based professional development. *Science Advances, 3*(2), e1600193.

Mendez, G.,Buskirk, T. D.,Lohr, S., et al. (2008). Factors associated with persistence in science and engineering majors: An exploratory study using classification trees and random forests. *Journal of Engineering Education, 97*(1), 57-70.

Michael, J. (2007). Faculty perceptions about barriers to active learning. *College Teaching, 55*(2), 42-47.

Moog, R. S.,Spencer, J. N. (2008). POGIL: An overview. In: ACS Publications.

Moog, R. S.,Spencer, J. N.,Straumanis, A. R. (2006). Process-oriented guided inquiry learning: POGIL and the POGIL project. *Metropolitan Universities, 17*(4), 41-52.

Ohland, M. W.,Sheppard, S. D.,Lichtenstein, G., et al. (2008). Persistence, engagement, and migration in engineering programs. *Journal of Engineering Education, 97*(3), 259-278.

Olson, S.,Riordan, D. G. (2012). Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics. Report to the President. *Executive Office of the President*.

Popova, M.,Shi, L.,Harshman, J., et al. (2020). Untangling a complex relationship: Teaching beliefs and instructional practices of assistant chemistry faculty at research-intensive institutions. *Chemistry Education Research and Practice, 21*(2), 513-527.

Rahman, T.,Lewis, S. E. (2020). Evaluating the evidence base for evidence‐based instructional practices in chemistry through meta‐analysis. *Journal of Research in Science Teaching, 57*(5), 765-793.

Raker, J. R.,Dood, A. J.,Srinivasan, S., et al. (2021). Pedagogies of engagement use in postsecondary chemistry education in the United States: results from a national survey. *Chemistry Education Research and Practice, 22*(1), 30-42.

Reinholz, D. L.,Apkarian, N. (2018). Four frames for systemic change in STEM departments. *International Journal of STEM Education, 5*(1), 1-10.

Sawada, D.,Piburn, M. D.,Judson, E., et al. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School science and mathematics, 102*(6), 245-253.

Seery, M. K. (2015). Flipped learning in higher education chemistry: emerging trends and potential directions. *Chemistry Education Research and Practice, 16*(4), 758-768.

Seymour, E.,Hewitt, N. M. (1997). *Talking about leaving*: Westview Press, Boulder, CO.

Seymour, E.,Hunter, A.-B.,Harper, R., et al. (2019). Talking about leaving revisited. *Talking About Leaving Revisited: Persistence, Relocation, and Loss in Undergraduate STEM Education*.

Shadle, S. E.,Marker, A.,Earl, B. (2017). Faculty drivers and barriers: laying the groundwork for undergraduate STEM education reform in academic departments. *International Journal of STEM Education, 4*(1), 1-13.

Simonson, S. R.,Earl, B.,Frary, M. (2021). Establishing a framework for assessing teaching effectiveness. *College Teaching*, 1-18.

Stains, M.,Harshman, J.,Barker, M. K., et al. (2018). Anatomy of STEM teaching in North American universities. *Science, 359*(6383), 1468-1470.

Stains, M.,Pilarz, M.,Chakraverty, D. (2015). Short and long-term impacts of the Cottrell scholars collaborative new faculty workshop. *Journal of Chemical Education, 92*(9), 1466-1476.

Stowe, R. L.,Cooper, M. M. (2019). Assessment in chemistry education. *Israel Journal of Chemistry, 59*(6-7), 598-607.

Strenta, A. C.,Elliott, R.,Adair, R., et al. (1994). Choosing and leaving science in highly selective institutions. *Research in higher education*, 513-547.

Sturtevant, H.,Wheeler, L. (2019). The STEM faculty instructional barriers and identity survey (FIBIS): Development and exploratory results. *International Journal of STEM Education, 6*(1), 1-22.

Sunal, D. W.,Hodges, J.,Sunal, C. S., et al. (2001). Teaching science in higher education: Faculty professional development and barriers to change. *School science and mathematics, 101*(5), 246-257.

Turpen, C.,Dancy, M.,Henderson, C. (2016). Perceived affordances and constraints regarding instructors' use of Peer Instruction: Implications for promoting instructional change. *Physical Review Physics Education Research, 12*(1), 010116.

Walczyk, J. J.,Ramsey, L. L. (2003). Use of learner‐centered instruction in college science and mathematics classrooms. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 40*(6), 566-584.

Walder, A. M. (2015). Obstacles to innovation: The fear of jeopardising a professorial career. *British Journal of Education, 3*(6), 1-16.

Walter, E.,Henderson, C.,Beach, A., et al. (2016). Introducing the Postsecondary Instructional Practices Survey (PIPS): A concise, interdisciplinary, and easy-to-score survey. *CBE—Life Sciences Education, 15*(4), ar53.

Wilson, S. B.,Varma-Nelson, P. (2016). Small groups, significant impact: A review of peer-led team learning research with implications for STEM education researchers and faculty. *Journal of Chemical Education, 93*(10), 1686-1702.

Woodbury, S.,Gess-Newsome, J. (2002). Overcoming the paradox of change without difference: A model of change in the arena of fundamental school reform. *Educational Policy, 16*(5), 763-782.

Yik, B.,Raker, J.,Apkarian, N., et al. (2022). Evaluating the impact of malleable factors on percent time lecturing in gateway chemistry, mathematics, and physics courses. *International Journal of STEM Education, 9*(1), 1-23.

CHAPTER 2. Development of the Departmental Climate around Teaching (DCaT) Survey: Neither psychological collective climate nor departmental collective climate predicts STEM faculty’s instructional practices

Chapter 2 was reproduced with permission from Springer Nature, and reprinted and modified with permission from *International Journal of STEM Education, **2021**, 8(1), 1-20.*

Introduction

A wave of instructional reforms within the last decade has focused on learner-centered instructional practices in Science, Technology, Engineering, and Mathematics (STEM) courses at the postsecondary level. However, uptake has not reached the desired level (Stains, et al., 2018) and efforts have been on-going to identify levels (e.g., national, institutional, departmental, individual) and levers (e.g., promotion and tenure guidelines, evaluation of teaching, professional development opportunities) that would increase the pace of uptake. Multiple studies have demonstrated the complexity of STEM instructors’ working environment (Anderson, Banerjee, Drennan et al., 2011; Austin, 2011; Brownell & Tanner, 2012; Childs, 2009; Froyd, 2011; Gess-Newsome, et al., 2003; C. Henderson & M. H. Dancy, 2007; Henderson & Dancy, 2011; Hora, 2012; T. J. Lund & Stains, 2015; Walczyk, Ramsey, & Zha, 2007) and highlighted the importance of taking a system-approach to instructional change (Austin, 2011; Elrod & Kezar, 2016). Departments have been recognized as a key level of the system to target and several recent efforts and frameworks aim to explore approaches to promote change at this level (Austin, 2011; J. C. Corbo, D. L. Reinholz, M. H. Dancy et al., 2016; Reinholz & Apkarian, 2018; Reinholz, Corbo, Dancy et al., 2017; Wieman, Perkins, & Gilbert, 2010). One characteristic of a department that has been advanced as critical to address is its climate around teaching, i.e., perceptions of policies, practices and expected behaviors related to teaching.

Several studies have found that STEM faculty point to departmental climate around teaching as a barrier to instructional change (e.g., (C. Henderson & M. H. Dancy, 2007; Shadle, et al., 2017; Sturtevant & Wheeler, 2019). For example, these three studies found that faculty cited departmental norms defined by lecture-focused teaching as a barrier to using learner-centered instructional practices. Interestingly, Shadle, Marker, and Earl (Shadle, et al., 2017) also surveyed STEM faculty at the authors' institution about drivers to instructional change. In alignment with other studies (Bouwma-Gearhart, 2012; Wieman, Deslauriers, & Gilley, 2013), they found departmental-level drivers for instructional change such as having discussions about teaching within the department and being encouraged to explore within their own teaching. An assumption implied by these results is that there can be an "inhibiting" or "supportive" departmental climate around teaching. However, no studies have operationalized these constructs. Moreover, although departmental climate has been advanced as a barrier to instructional change, few studies have explored the relationship between adoption of learner-centered instructional practices and departmental climate around teaching (Bathgate, Aragon, Cavanagh et al., 2019; Borda, Schumacher, Hanley et al., 2020; T. J. Lund & Stains, 2015).

Several instruments have been developed to measure the climate of an organization and could thus be leveraged to design an instrument focused on departmental climate around teaching (Table 2.1). However, only five out of the eleven instruments in Table 2.1 were developed for the higher education setting and within this sample, few focused on the department as the organization. Other instruments measure the climate of K-12 and industry organizations. Moreover, a review of the studies designing and employing these instruments pointed to methodological and analytical shortcomings (Patterson, West, Shackleton et al., 2005). First, most climate surveys lack the validity measures recommended by the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, National Council on Measurement in Education et al., 1999). As Table 2.1 indicates, 42% of the reviewed instruments did not provide evidence of content test and/or internal structure; none of the instruments included cognitive interviews to provide evidence of the type of validity known as response processes. There is thus a need to design a climate survey following the recognized standards for instrument development. Second, studies that focused on measuring climate around teaching within an organization (i.e., school, institution, or department) characterized individual faculty's description of their colleagues' perceptions of the climate within their organization, which is known as the psychological collective climate. However, the industry-focused and K-12 climate literature measures the extent to which there is a consensus among the respondents of the organization; this is known as the organizational collective climate (Ostroff, 1993; Schneider, Ehrhart, & Macey, 2013). In other words, studies need to statistically investigate whether faculty within a department are answering a climate survey in a similar way and therefore have a consensus view on the climate. Such measurement of departmental consensus is needed in order to substantiate claims regarding the importance or lack thereof of departmental collective climate on teaching in relation to the adoption of learner-centered instructional practices. No studies that we could find on departmental climate around teaching attempted to measure departmental collective climate.

Table 2.1 List of surveys measuring climate within an organization.

Year	Name of survey	Author(s)	Type of organization	Description of instrument development	Content test	Response process	Internal structure	Relations of other variables
1963	Organizational Climate Description Questionnaire	Halpin &Croft	Elementary school	No	No	No	No	No
1967	Relationship of Centralization to Other Structural Properties	Hage & Aiken	Social welfare and health agencies	Yes	Yes	No	No	Yes
1993	Climate Survey	Ostroff	Secondary school	Yes	No	No	No	Yes
1997	The Organizational Climate Questionnaire	Furnham.& Goodstein	Industry	Yes	Yes	No	No	Yes
2005	Organizational Climate Measure	Patterson et al	Manufacturing sector	Yes	Yes	No	Yes	Yes
2007 **2010**	Higher Education Research Institute Faculty survey	Lee Hurtado	Higher education	Yes	No	No	Yes	Yes
2009	Organizational Change Questionnaire	Bouckenooghe, Devos & Van den Broeck	Public and private sector organizations	Yes	Yes	No	Yes	No
2012	Department Teaching Climate	Knorek	Higher education	Yes	Yes	No	Yes	Yes
2014	Survey of Climate for Instructional Improvement	Walter, Beach, Henderson & Williams	Higher education	Yes	Yes	No	Yes	Yes
2017	Instructional climate survey	Landrum, Viskupic, Shadle & Bullock	Higher education	Yes	Yes	No	Yes	No
2018	Collaborative on Academic Careers in Higher Education	Mamiseishvili, & Lee	Higher education	Yes	Yes	No	Yes	Yes

This study is the first empirical exploration testing the extent to which psychological collective climate and departmental collective climate around teaching predict the level of uptake of learner-centered instructional practices. We developed and collected the Departmental Climate around Teaching (DCaT) survey from STEM faculty at four-year institutions of higher education in the United States to explore the following research questions:

1. What validity and reliability evidence support the use of the Departmental Climate around Teaching (DCaT) survey?
2. To what extent do different types of psychological collective climate around teaching exist within STEM departments?
3. To what extent does a departmental collective climate around teaching exist at the departmental level?
4. To what extent does the psychological collective climate around teaching and departmental collective climate around teaching relate to the use of learner-centered instructional practices within a department?

Conceptual Framework

Organizational culture versus organizational climate

Organizational culture and organizational climate are two terms employed to describe people's experiences in the settings of their working environment (Schneider, et al., 2013). These terms are often misused in the literature (e.g., studies focused on culture mistakenly measure climate and vice versa) and thus it is essential to understand the differences between these two terms.

Organizational culture is defined as "the shared basic assumptions, values, and beliefs that characterize a setting and are taught to newcomers as the proper way to think and feel, communicated by the myths and stories people tell about how the organization came to be the way it is as it solved problems associated with external adaptation and internal integration" (Schneider, et al., 2013).

Organizational culture applied to the academic department represents for example the way faculty members follow unwritten rules when making departmental decisions for promotion and tenure. Culture is typically studied through qualitative case studies.

Organizational climate refers to "the shared perceptions of and the meaning attached to the policies, practices, and procedures employees experience and the behaviors they observe getting rewarded and that are supported and expected" (Schneider, et al., 2013). Organizational climate applied to the academic department represents the faculty members' current perceptions and attitudes of how teaching is evaluated. Research focused on organizational climate typically relies on surveys; data is analyzed in an aggregated way by using different statistical methods and different consensus models. For the remainder of this study, we will be focusing on the latter, i.e., organizational climate, with academic department as the organization.

Two measures of organizational climate

Approaches to measure climate within an organization differ between studies in Discipline-Based Education Research (DBER) and studies in K-12 and industry settings. In both fields, researchers often ask each organizational member how they think others in the organization perceive organizational policies, practices, and procedures (Ostroff, 1993; Patterson, et al., 2005). However, in the K-12 and industry studies, researchers then leverage statistical methods to explore the extent to which there is a consensus on these perceptions among the members of the organization (Ehrhart & Schneider, 2016). If consensus is demonstrated, they aggregate the individual level data into a single climate measure. At this point, the measure represents the collective view of the organizational climate, i.e., the climate at the organization level (e.g., school) and not the individual level (e.g., teacher). This aggregated measure is then used to test whether the organizational collective climate has an impact on a desired outcome for this type of organization (e.g., productivity, job satisfaction). In contrast, studies on institutional or departmental climate in DBER aggregate the perceptions collected from individual members (e.g., take an average of the perceptions across members of a department) without testing for consensus across respondents (Landrum, Viskupic, Shadle et al., 2017; Ngai, Pilgrim, Reinholz et al., 2020). The aggregated value in these studies represents the average perception of the organizational climate by members of the organization (i.e., individual level measure of climate of the institution or department) and not the collective perception of the organizational climate (i.e., departmental or institutional level measure of climate).

The distinction in the analytical approaches employed to measure organizational climate between the K-12/industry literature and the DBER literature is important when studies are interested in exploring the relationship between organizational climate and other organizational outcomes. Although the two types of literature refer to what seems to be the same construct, organizational climate, DBER studies describe average individual perceptions of the climate while the organizational climate literature describes the collective perception of the climate. This different level in measuring organizational climate impacts claims that can be made about the relationship between such climate on organizational outcomes or characteristics such as uptake of learner-centered instructional practices.

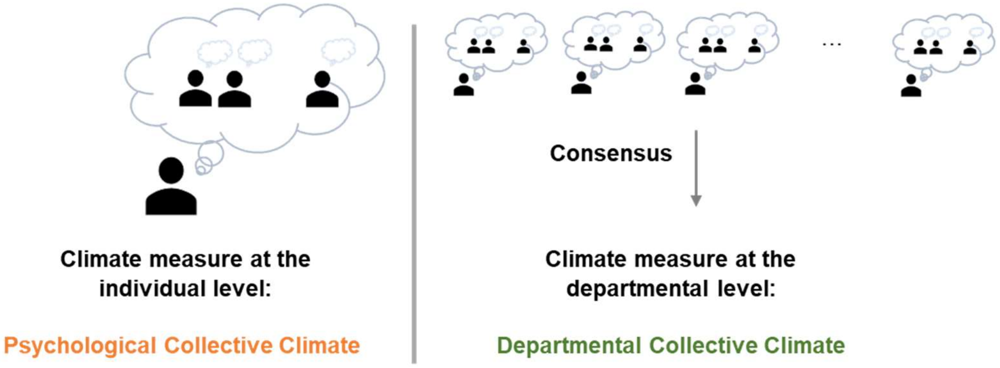
Chan (1998) highlighted the importance of explicitly identifying the level at which the organizational climate construct is being measured. He identified five composition models to assist researchers in developing a common framework to explain the relationship between constructs and level of the organization: “Composition models specify the functional relationships among phenomena or constructs at different levels of analysis (e.g., individual level, team level, organizational level) that reference essentially the same content but that are qualitatively different at different levels” (Chan, 1998, p. 234). Making the transformation of a construct across levels of the organization clear provides “conceptual precision in the target construct” (Chan, 1998).

Of the five models described by Chan, the approach followed by the K-12/industry literature matches the Referent-Shift Consensus Model (Ehrhart & Schneider, 2016). Applied to this study, the measure in the Referent-Shift Consensus Model describing the climate at the department-level represents the shared perceptions of the climate among the members of the department. This is typically done by calculating within-group agreement indexes (e.g., r_{wg}, (O'Neill, 2017)) for each factor or item in the survey and only aggregating the factors or items that meet a certain within-group agreement threshold. Moreover, the Referent-Shift Consensus Model shifts the focus from a member's personal perception of the climate in the organization to the member's thinking about how other members in the organization perceive the climate. For example, the following survey item "I embrace other colleagues' innovative teaching practices" would be rephrased "Overall, instructors in my department embrace other colleagues' innovative teaching practices." The model thus leads to two measures of climate (Figure 2.1):

1) "Psychological collective climate, is defined as the individual's description of other organizational members' perceptions of the climate" (Chan, 1998); This measure thus focuses on the individual members of the organization and their views on how others in their organization think of the climate; this is a measure of the climate of the organization at the individual level.

2) Organizational collective climate is a measure of the climate at the organizational level; this measure derives from the aggregation of measures collected at the individual level (i.e., psychological collective climate); aggregation is only justified when consensus and agreement among individuals' psychological collective climate perceptions have been demonstrated. Given that the definition of organizational climate is the shared perceptions of policies and practices and thus a property of the organization (Ehrhart & Schneider, 2016), organizational collective climate is the only measure out of the two that represents that construct.

The DBER literature has been measuring psychological collective climate. However, the analytical approach typically employed in these studies does not follow the Referent-Shift Consensus Model and rather matches a different model identified by Chan: the Additive Model. In the Additive Model, the measure at the department-level is the summation or average of measures collected at the faculty level. However, the variance among answers provided by the faculty is not considered relevant during this transformation. These studies thus do not measure shared perceptions of the climate, i.e., organizational (e.g., departmental) collective climate.



Consensus
Climate measure at the individual level:
Psychological Collective Climate
Climate measure at the departmental level:
Departmental Collective Climate

Figure 2.1 Difference between psychological collective climate and departmental collective climate

At the time of writing, it is unclear whether psychological collective climate and/or organizational collective climate is a predictor of instructional innovation within a department. In this study, we measured both the psychological collective climate and departmental collective climate through a new survey instrument that underwent rigorous validity and reliability studies. We then tested the extent to which each measure of climate predicted the level of use of learner-centered instructional practices within the department.

Methods

The survey employed in this study consisted of three parts: 1) the Departmental Climate Around Teaching (DCaT) survey (Appendix B3); 2) an abbreviated version of the Measurement Instrument for Scientific Teaching (MIST; (Durham, Knight, & Couch, 2017)) (Appendix B5), and 3) a demographic section (Appendix B4). All survey components are presented in the appendix. Participants spent an average of twenty minutes to answer this three-part survey, which was collected online via Qualtrics. All stages of the study were approved by the University of Nebraska-Lincoln Institutional Review Board.

Participants

To measure departmental collective climate, it is important to collect data from a large proportion of the faculty members in the department. Moreover, to address the fourth research question - relationship between climate and uptake of learner-centered instructional practices - it is necessary to capture a diverse set of instructional practices among the population surveyed. We devised two different strategies to maximize the likelihood of achieving these two goals. First, we selected STEM departments that had taken part in some department or institution-level change initiative related to teaching. Our goal was to identify settings where there may be some buy-in in assessing climate around teaching. We read 48 abstracts of grants funded by the National Science Foundation under the WIDER program (National Science Foundation, 2013) and identified eleven projects indicating change in departmental or institutional climate around teaching as one of their goals. We contacted the principal investigators of these projects and probed their interest in implementing our survey with their local population. One of the principal investigators helped us identify two departments at their institution that were likely to provide a high participation rate. Second, we selected departments with faculty who had participated in national professional development programs around teaching, specifically the Cottrell Scholar Collaborative New Faculty Workshop (Cottrell Scholars Collaborative, 2017) and Process Oriented Guided Inquiry (POGIL Team, 2019). For the CSC NFW, we selected departments in which at least three of the faculty had participated in the workshop. For the POGIL programs, we contacted the program leadership to help us identify departments in which POGIL was consistently implemented by at least one of the faculty.

In total, 727 emails were sent to faculty members representing 21 different institutions. We emailed all faculty members across ranks (e.g., lecturer, tenure-track faculty) within each of the 22 departments. In total, 201 instructors completed the survey, which corresponds to a raw response rate of 28%. Raw response rates by department are provided in Appendix C2. The data set was then cleaned by deleting 1) participants who had no teaching responsibility, 2) participants who did not answer all items; 3) participants who used “I don’t know” option for any of the items in the DCaT survey (Cole, 1987), 4) items for which more than 5% of the participants answered “I don‘t know” (Cole, 1987). The cleaned sample size was 166 instructors which corresponds to a 23% response rate. Twenty-two departments are included in the sample, most of them being chemistry departments (Appendix C2).

Development of the Departmental Climate around Teaching (DCaT) survey

We followed guidelines from the Standards for Educational and Psychological Testing to develop the DCaT survey. In particular, we abided by the following steps to test the extent to which the survey provided valid and reliable data about psychological collective climate: 1) test content, 2) response processes, 3) internal structure, and 4) internal consistency.

Test Content

Test content is used to test whether an instrument is capturing the intended domain. In our case, it was the psychological collective climate (American Educational Research Association, et al., 1999). Test content is typically established through consultation with experts in the target domain. We conducted a literature review to identify typical constructs considered by studies measuring organizational climate. This review focused on studies describing the development and/or use of surveys measuring organizational climate and was not limited to educational settings since this body of work is minimal. Studies were selected based on whether they 1) measured climate around teaching (Knorek, 2012; Landrum, et al., 2017; E. Walter, Beach, Henderson et al., 2014), 2) measured institutional or organizational climate from an instructor’s perspective (Eagan, Stolzenberg, Lozano et al., 2014; Halpin & Croft, 1963; Mamiseishvili & Lee, 2018; Ostroff, 1993), or 3) described popular instruments designed to measure organizational collective climate outside of an education setting (Furnham & Goodstein, 1997; Hage & Aiken, 1967; Patterson, et al., 2005; Thomas & Tymon Jr, 2009). Eleven studies met these criteria (Table 2.1) and were used to identify common constructs assessed across these studies as these would indicate a certain level of saliency when describing an organizational climate. We then leveraged items from these surveys to develop the first version of our survey. This initial draft (DCaT Version 1; Appendix B1) was then shared with DBER experts for their feedback. This feedback was used to modify the survey (DCaT Version 2; see Results section).

Response Process

This step focuses on testing whether the survey items and methods used to collect answers within the survey (e.g., type of Likert scale) are interpreted by the study participants the way the developers of the survey had intended. Testing for response processes is typically conducted via cognitive interviews (Peterson, Peterson, & Powell, 2017). We gathered a convenient sample of 11 faculty members from seven institutions in the United States to participate in the cognitive interviews (Robinson, 2013). Three of the faculty members worked at primarily undergraduate institutions, while the other eight worked at research-intensive institutions. The academic appointments of the participants included: two full professors, six associate professors, one assistant professor, one associate professor of practice, and one lecturer. Nine were chemists and two were biologists. The diverse characteristics of the interview participants support the use of the survey for instructors at different academic ranks and from different types of institutions. The interviews were conducted once with each faculty in four rounds (about three faculty per round) with revisions of the survey between each round. All interviews were conducted within a three-month period. The cognitive interviews were conducted via the Zoom online platform or in a private room to ensure the confidentiality of the participants.

Participants were first asked to fill out the DCaT Version 2 and demographic portion of the survey. Within a week following completion of the survey, we interviewed them. Interviews typically lasted thirty minutes and engaged the faculty members in a think-aloud process to identify their understanding of items and rationale for choosing a response option for an item, to unpack inconsistencies within their responses, and to collect their feedback on the survey design (interview protocol see Appendix A1). The cognitive interviews led to further refinement of the survey (see Results section). This DCaT Version 3 (Appendix B2) was used to test the internal structure of the survey.

Internal Structure

Establishing the internal structure of an instrument consists of providing evidence about the relationships between items and the constructs intended to be measured. The third version of the DCaT survey along with the adapted MIST (Appendix B5) and demographic section (Appendix B4) were embedded in Qualtrics as an online instrument. Confirmatory Factor Analysis (CFA) was employed to validate the internal structure (Arjoon, Xu, & Lewis, 2013) of the survey. Due to the categorical nature of the Likert scale responses, the Weighted Least Square Mean and Variance adjusted (WLSMV) estimator was used to conduct the CFA in Mplus version 7.4. A minimum recommended sample size to conduct CFA is five to ten respondents per item (T. A. Brown, 2014). In this study, the sample size met the minimum criteria of five participants per item. Since we did not expect certain items within a construct to be more important than others in describing the construct, we followed the guidelines provided by Komperda, Pentecost, and Barbera (2018) and used the Tau-equivalent indicator. In this process, all factor loadings within the same factor are set to be identical to ensure that each item within the same construct is equally weighted. This allowed us to use the mean across items within a factor to describe the factor.

The model fit was tested by exploring two typical model fit indexes for CFA analyses that are independent of sample size: the comparative fit index (CFI) indicates an adequate fit when its value is above 0.9 (Hu & Bentler, 1999) (range from 0 to 1); The root mean square error of approximation (RMSEA), with values less than 0.08 indicating a good fit (Browne & Cudeck, 1993).

Internal Consistency

Testing for internal consistency provides evidence that respondents to an instrument answer similar items in similar ways. The internal consistency of the survey was measured with Cronbach’s alpha coefficient since it has been suggested to be an acceptable reliability measure for Tau Equivalent models (Komperda, et al., 2018). A cut-off value of >0.7 is recommended (Komperda, et al., 2018; Streiner, 2003).

Abbreviated version of the Measurement Instrument for Scientific Teaching (MIST)

The literature on instructional change indicates that faculty members often point to their departmental climate as a barrier to using learner-centered instructional practices. The level of use of these practices was measured using an abbreviated version of the Measurement Instrument for Scientific Teaching (MIST, Appendix B4; (Durham, et al., 2017)). This instrument was chosen among many others because it underwent the most rigorous validation and reliability studies (Durham, Knight, Bremers et al., 2018). MIST measures the extent to which STEM faculty employ Scientific Teaching practices in their classroom through self-report from faculty. The instrument measure eight subcategories of Scientific Teaching practices (Active-learning strategies, learning goals use and feedback, inclusivity, responsiveness to students, experimental design and communication, data analysis and interpretation, cognitive skills, and reflection). While all these categories fall under learner-centered instructional practices, in this study we used only the Active-learning strategies subcategory since it has been demonstrated to be the MIST subcategory that correlates the strongest with students' and external observers' reports of the level of active learning in a course (Durham, et al., 2018). Considering our expected modest sample size to establish internal structure, we consulted with the MIST developers to identify a sub-set of five items within that Active-learning strategies subcategory that statistically correlated best with external measures of presence of active learning. These five items include: 1) average percent of class time during which students were asked to answer questions, solve problems, or complete activities other than listening to a lecture; 2) frequency of use of polling methods; 3) frequency of use of in-class activities other than polling methods; and frequency of 4) in and 5) out-of-class group work activities.

Confirmatory factor analysis for this abbreviated version of MIST was conducted using Mplus version 7.4 with a Robust Maximum Likelihood (MLR) estimator and congeneric indicator. The

Comparative Fit Index (CFI) for all five items was 0.981, and the Standardized Root Mean Square Residual (SRMR) value was 0.039. However, the factor loadings for two items (frequency of use of polling methods and frequency of group work for outside the lecture hall) were lower than 0.2 (Sharma, Mukherjee, Kumar et al., 2005) and thus deleted. Since there were only three items left to describe this one factor, a large sample size was needed (more than 300) to conduct the CFA (MacCallum, Widaman, & Sehee Hong, 1999). Our sample size (*N*= 166) did not meet this criterion, so we relied on a reliability test. Cronbach's alpha for the abbreviated version of MIST was 0.80, which indicated that these three items were answered in a consistent manner by participants.

A MIST scale score was generated following the equation given in the original paper for each participant (Durham et al., 2017):

$$\text{MIST}_{\text{scale score}} = [(XQ1+XQ2+\ldots+XQn)/n] \qquad \textit{Eq. 2.1}$$

where XQ1… XQn are the normalized response for each question, and n is the number of items included in the scale score calculation. The MIST score was used as a dependent variable in a simple linear regression analysis.

Analysis

Mixture model cluster analysis

Cluster analysis was used to identify groups of faculty members who provided similar psychological collective climate descriptions. Mixture model clustering (MMC) analysis was used in this research for the following reasons:

1) The MMC method, unlike the heuristics methods of clustering such as K-means is able to make population-wide estimations (Landau & Chis Ster, 2010);

2) The heuristic methods require a preassigned number of clusters before the analysis while the MMC method provides empirical recommendations for the number of clusters (Fraley & Raftery, 1998);
3) The MMC method allows for comparison of different solutions by using fit indices.

Bayesian Information Criterion (BIC) was used to identify the class enumeration.

Regression analysis to explore the relationship between climate and instructors' uptake of learner-centered instructional practices.

A simple linear regression was used to predict instructors' use of learner-centered instructional practices based on types of psychological collective climate. Since the type of psychological collective climate was categorical in nature, dummy coding was used to recode this variable.

$$y_i = \beta_0 + \beta_1 X_1 + \ldots \beta_i X_i \qquad Eq.\ 2.2$$

y_i is the dependent variable which refers to the abbreviated MIST scores in this study, X through X_i are different clusters identified in the MMC analysis which served as independent variables, β_0 is the constant in the regression equation, and β_1 through β_i are standardized coefficients for different clusters.

Measuring the departmental collective climate

The measure of departmental collective climate necessitates a high response rate. This measure was calculated for the three departments that initially had at least half of their instructors answer the DCaT survey (see raw response rate in Appendix C2). The cleaned response rates for these three departments are provided in Appendix C3.

Inter-rater agreement, $r_{wg(J)}$, which indicates the level of agreement among faculty members within the same department, was calculated using the following equation:

$$r_{wg(J)} = \frac{\mathrm{J}\,(1 - \frac{\bar{s}_x^2}{\bar{s}_{EU}^2})}{\mathrm{J}\left(1 - \frac{\bar{s}_x^2}{\bar{s}_{EU}^2}\right) + \frac{\bar{s}_x^2}{\bar{s}_{EU}^2}} \qquad Eq. 2.3$$

where, J is the number of items in the construct, $\bar{s}_x^2$ is the obtained average variance of the items in the construct, and $\bar{s}_{EU}^2$ is the variance of the uniform distribution. A value of $r_{wg(J)}$ above 0.75 indicates a high level of within-group inter-rater agreement (O'Neill, 2017).

Intra-class correlation coefficients (ICC) do not investigate absolute agreement but consistency between faculty members within a department. Two coefficients were calculated: ICC(1) and ICC(2). ICC(1) indicates whether the mean of one faculty is a reliable measure of the mean of another faculty within the same department. ICC(2) indicates whether there is a reliable difference between means across different departments (Koo & Li, 2016). The equations for calculating ICC(1) and ICC(2) are listed below:

$$\mathrm{ICC(1)} = \frac{\mathrm{BMS} - \mathrm{WMS}}{\mathrm{BMS} + (\mathrm{k} - 1)\mathrm{WMS}} \qquad Eq. 2.4$$

$$\mathrm{ICC(2)} = \frac{\mathrm{BMS} - \mathrm{WMS}}{\mathrm{BMS}} \qquad Eq. 2.5$$

where BMS is the between-treatments mean square and WMS is the within-treatment mean square.

Results

Validity and reliability evidence supporting the use of the Departmental Climate around Teaching survey
Modifications based on the test content study

Constructs and items were initially identified through a literature review of studies measuring teaching climate or organizational climate within an educational setting as well as highly cited studies describing a survey measuring organizational collective climate in organizations outside academia. Twelve studies were identified through this review process (Appendix C1) contains a list of the studies and the constructs measured in each). We considered constructs that had been included in at least four of these studies, which included: Formalization, Cooperation, Participation, Supervisor Support, Warmth, Growth, Innovation, Autonomy, Achievement, Extrinsic Reward, Performance Feedback, and Outward Focus. One of these constructs was not included in the first version of the DCaT survey: Formalization - i.e., “perception of formality and constraints in the organization; emphasis on rules, regulations and procedures” (Ostroff, 1993, p. 62) - was a construct found in K-12 and industry-focused studies and was not considered relevant when exploring the climate around teaching in the higher education setting. We added the construct of Resources since funding, space, and teaching budget have been identified as barriers to uptake of learner-centered instructional practices in prior studies (Sturtevant & Wheeler, 2019) and this construct was included in the most recent climate surveys developed for STEM higher education contexts (Landrum, et al., 2017; E. Walter, et al., 2014). The first draft of the survey (DCaT Version 1, Appendix B1), which contained 48 items was developed based on this analysis. 36 items came from these prior studies (as is or modified if needed to adapt to context) and 12 items were developed by the authors. This draft was shared with the University of Nebraska-Lincoln DBER group during Spring 2018 to collect their feedback. The survey was further revised as a result of this consultation (DCaT Version 2). For example, the DBER group pointed out that the targeted population was not clearly described. Throughout the survey, we had used the word “instructors” but it was unclear to the DBER group who should be included as an instructor. For example, certain STEM departments heavily rely on graduate teaching assistants to teach lectures, laboratories, or recitations. To clarify our targeted population, we added the following description at the beginning of the survey: "Instructors" in this questionnaire refer to faculty who teach undergraduate level courses (including lecturer, tenured/tenure-track professor, professor of practice, but EXCLUDING graduate students)." Graduate teaching assistants were not included as instructors since they have limited involvement in decisions related to teaching and curriculum conversations within a department (e.g., they are not assigned to or made aware of decisions made during curriculum committees).

Modifications based on the response process study

The revised version of the survey (DCaT Version 2) was then distributed to faculty members who had agreed to take part in cognitive interviews. Faculty answered the survey prior to being interviewed (see the method section for details on the response process procedures). The cognitive interviews helped us to identify several critical issues with this version of the instrument.

The first issue was that the participants did not consider the departmental climate as a shared perception of all instructors within their department when they answered the questions. Items in several of the surveys identified in the literature started with "Instructors in my department." We either used these items or developed items that used the same wording. However, interviewees highlighted that this could lead to different interpretations as the following quote exemplifies: "*When answering questions, you use frequency, others may use your own data points. Some items, you may think about a certain meeting"* (Participant #4). We thus changed each item by adding at the beginning the word "overall." This change was made halfway through the collection of the interviews. Participants interviewed afterwards demonstrated an understanding that instructors should be considered as a collective when answering the survey.

Second, it became clear during the interviews that the response option provided – a 5-point Likert scale going from strongly agree to strongly disagree - did not clearly capture participants' opinions, especially when choosing the neutral option (i.e., neither agree nor disagree): *“I can’t strongly answer that. Some of them [items], like I wonder if you ... so you don't have an N/A or can't answer category, right? Which, so I kind of was using that middle one [neither agree nor disagree] as that”* (Participant #4). We had debated as to whether the neutral position should be included in the first place since many of the surveys we were consulting did not include it. However, the interviews confirmed to us that it was needed and that another option “I don’t know” should be added. Including these two extra response options helped respondents feel that they were not forced to agree or disagree when they were truly on the fence or did not have enough information to decide. It also helped us ensure that the neutral option was interpreted consistently across the participants.

Third, most of the climate surveys do not ask the participants to answer the survey based on a particular time frame. This issue was raised during the interviews as the following quote illustrates: *"Some questions need a time frame. Past five years? Or past three years? The answer will be different*" (Participants #9). For example, a faculty member who has been in their department for 20 years may have experienced different climates during their time and may try to think across all of those when responding to the survey. Since prior studies have shown that departmental leaderships can influence instructors' perceptions of the teaching context within the department (Ramsden et al., 2007), we decided to align the time frame with the period of activity of the current department chair. We thus added the following question prior to answering the DCaT survey (Table S4): "How long has your chair been in his/her current position?" We then added the following sentence to the instruction for the DCaT survey, leveraging the 'piped text' option in Qualtrics (Table S3): "When answering the survey, please focus on the last '*piped text from the chair question*' year(s)." We purposefully did not refer to the chair in the instruction so as to limit potential (conscious or unconscious) influence that it could have on respondents.

Fourth, several constructs typically included in climate surveys were eliminated as a result of the analysis of the cognitive interviews: Extrinsic Rewards, Resources, and Warmth. Interviewees indicated having trouble answering items associated to the Extrinsic Rewards construct for several reasons. One of the faculty from a primary undergraduate institution indicated that the evaluation of teaching did not occur at the department-level but rather at the provost level. Moreover, hiring processes at their institution were conducted at the college-level not the department. Therefore, reward for teaching excellence in the form of hiring or promotion was not controlled by the department, making the extrinsic award items irrelevant for this participant. Another participant from a primary undergraduate institution indicated that evaluation of teaching was conducted at the Chair and Dean's levels. We assume that these situations may not be unique among primary undergraduate institutions. One of the items for extrinsic rewards focused on teaching awards (see Table S1). One of the participants at a research-intensive institution indicated that factors other than teaching excellence may come into play for teaching award nominations:

> *"People are nominated for awards a lot, but that doesn't mean that you need to be a really great teacher to continue in your job. […] The department nominates for teaching awards every year because we're not just going to sit out on a cycle where we're going to nominate people for awards no matter what." Participant #1*

Moreover, four participants indicated that there are no respected, highly desirable awards for teaching like there are for research.

> *"So you don't get pay raises for your teaching, don't get recruited. You don't get counteroffers. The administration, they pass around some awards and what are those awards giving you? There's $0 million behind that award*." Participant #4

> *"We know that the money follows the research awards, but doesn't follow the teaching awards."* Participant #5

Similarly, four of the eleven interview participants indicated that there was no good measure of effective teaching making it difficult to reward and recognize it.

> *"We place a high premium on quality, but we have no metrics to assess that. So if one, if a chair think you're a good teacher, then you're a good teacher"* Participant #3

> *"We don't have any sort of standardized measure of teaching performance. […] There are actually very few sort of strong public indicators of teaching achievements."* Participant #8

Consequently, participants found it challenging to answer the item "Overall instructors in my department carefully consider evidence of effective teaching when making decisions about continued employment and/or promotion."

The Resources construct captures the tools used for instructional improvement, including time, funding, office space, equipment, and support services. Our interviewees indicated that most of these resources were not controlled by the department and thus did not feel it was appropriate to ask these questions: ***"****Time as a resource is not controlled by the department"* and *"Resource is provided at the college level, not at the departmental level."* The Warmth construct, which captures whether informal communications occur among faculty members, was eliminated because faculty members indicated that they did not have enough information to answer this type of questions. Here are some quotes to illustrate this point:

> *"That's actually really hard to get these questions, unless you know everyone very well informally and attend these happy hours or lunches and things like that, which I do not know if many, I don't know the answer."* Participant #9
>
> *"Oh yeah. Yeah. I have no clue. So, in that case, I don't know what my colleagues are doing and so I do not know."* Participant #10

Fifth, the Participation and Cooperation constructs were combined. The construct Participation was defined as measuring whether faculty members were involved in decision-making process and setting goals/policies with respect to the teaching mission of the department. Two of the items were removed since the cognitive interviews illustrated that adoption of teaching methods does not occur at the departmental level: *"I don't know if maybe this, all right, so we don't as a department make decisions on new teaching methods"* (Participants #4). Regarding the construct Cooperation, two of the original items were moved to "supervisor support" which is more focused on capturing the helpfulness of the departmental leadership. Since both constructs Participation and Cooperation had only two items each and were conceptually similar, we combined them into one construct, Involvement. We measured the internal consistency of the four items by using Cronbach's alpha, and obtained a 0.85 indicating that the four items were measuring similar ideas. We named the new construct Involvement which we defined as faculty members' perceptions of involvement in decision-making processes, setting goals, policies with respect to the teaching mission of the department, and the willingness to communicate about teaching related issue with colleagues.

Following the cognitive interviews, the third version of the DCaT survey (Appendix B2) consisted of eight constructs (32 items): Involvement, Growth, Autonomy, Supervisor Support, Innovation, Outward Focus, Achievement, and Performance Feedback. Each construct's definition is provided in Table 2.2. A six-point Likert scale was used as the item response options from "strongly disagree"(as 1) to "strongly agree" (as 5) with the sixth point represented by "I don't know."

Table 2.2 Definitions of the constructs measured in the Departmental Climate around Teaching survey

Construct	Definition
Involvement	Perception that faculty are involved in decision-making processes and setting goals and policies with respect to the teaching mission of the department.
Growth	Perception of emphasis on personal development with respect to teaching.
Supervisor Support	The extent to which instructors experience support and understanding from their department chair.
Autonomy	Perception of having the freedom to plan and control over their own curriculum.
Innovation	Perception of emphasis on innovation and creativity in teaching approaches and accepting of changes in instructional practices.
Outward focus	The extent to which the department is responsive to the needs of students.
Achievement	Perception of challenges, demand for work, and continuous improvement of performance.
Performance feedback	The measurement of and feedback on teaching performance.

Internal structure of the DCaT survey

Before evaluating the internal structure of the DCaT survey, the data set was cleaned up (see methods section). Items for which more than 5% of the participants answered "I don't know" were deleted (supervisor support-1 and achievement-3). Consequently, 30 items were used to explore the internal structure of the DCaT survey.

Table 2.3 Standardized factor loadings from the confirmatory factor analysis of the DCaT survey (*n*=166)

Factor	**Item**	**Factor loading**
Involvement	Invo1	0.84
	Invo2	0.84
	Invo3	0.84
	Invo4	0.84
Growth	Grow1	0.79
	Grow2	0.79
	Grow3	0.79
	Grow4	0.79
Supervisor Support	Supp2	0.88
	Supp3	0.88
	Supp4	0.88
Autonomy	Auto1	0.84
	Auto2	0.84
	Auto3	0.84
	Auto4	0.84
Innovation	Inno1	0.85
	Inno2	0.85
	Inno3	0.85
	Inno4	0.85
Outward focus	Outw1	0.77
	Outw2	0.77
	Outw3	0.77
	Outw4	0.77
Achievement	Achi1	0.88
	Achi2	0.88
	Achi4	0.88
Performance feedback	Feed1	0.93
	Feed2	0.93
	Feed3	0.93

The Comparative Fit Index (CFI) for all 30 items was 0.654 (below the 0.9 threshold) indicating a lack of fit. We consulted the Modification Indices (M.I.) in order to identify items that could improve the model fit (Ab Hamid, Mustafa, Idris et al., 2011). We considered whether to eliminate items with high M.I. values based on their alignment with the construct they were intended to measure. As a result, the item performance feedback-4 was deleted. Indeed, this item had high M.I. value and was focused on the measures used to evaluate teaching rather than the feedback provided to faculty and was thus not as well aligned with the construct as the other three items. Another CFA analysis was conducted on the 29 items (Table 2.3).

The CFI for this model was 0.946 (above the 0.9 threshold) and the Root Mean Square Error of Approximation (RMSEA) was 0.076 (below the 0.080 threshold). These results thus indicated that this final version of the DCaT survey (DCaT Version 4; Appendix B3) met the goodness-of-fit standards.

Internal consistency

Internal consistency of each construct was evaluated with Cronbach's alpha (Table 2.4). We observed values ranging from 0.77 to 0.93, which indicated that the data from the DCaT Version 4 were sufficiently reliable for interpretation.

Table 2.4 Cronbach's alphas, means, and standard deviations for each construct measured in the DCaT survey (*n*=166)

Constructs	Cronbach's α	Mean	SD
Involvement	0.83	3.75	0.85
Growth	0.78	3.11	0.80
Autonomy	0.82	3.95	0.78
Supervisor Support	0.87	3.67	0.94
Innovation	0.88	3.43	0.78
Outward Focus	0.82	3.56	0.84
Achievement	0.85	3.95	0.73
Performance feedback	0.83	2.87	1.00

Types of Psychological Collective Climate around Teaching in STEM departments

Once it was demonstrated that the DCaT survey provided valid and reliable data, we endeavored to answer the second research question: To what extent do different types of psychological collective climate around teaching exist within STEM departments?

Overall, faculty members in this study reported a neutral to positive psychological collective climate around teaching in their department as indicated by the range of means for each construct (from 2.87±1.00 to 3.95±0.73 on a scale from 1-5, 5 being strongly agree; Table 2.4). We conducted a Mixture Model Clustering analysis to identify groups of faculty members across the whole sample who provided similar psychological collective climate descriptions. To identify the model that best fit the data, we examined the Bayesian Information Criterion (BIC) and the BIC difference (Table 2.5). Although the "VEE" model had the lowest BIC value, the sample within this model was evenly split between the two clusters and the interpretation of the clusters lacked nuance, i.e., estimated means were similar across the two clusters for most of the constructs (Appendix C4). The next best model, "VEI", had a four-cluster solution, which provided an interpretable and nuanced description of the types of climate perceived by the sample. Consequently, VEI was selected for the rest of the analysis.

Table 2.5 Mixture Model Clustering (MMC) Class Enumeration (*n*=166)

Model type, Number of class	**VEE, 2**	**VEI, 4**	**VEI, 3**
BIC	-2930.7	-2945.5	-2946.4
BIC difference	0.0	-14.8	-15.7

The estimated model proportion of each of the four types of climate are listed in Table 2.6 along with the estimated means for each construct. We leveraged results presented in Table 6 to describe each climate. Of note, the construct Autonomy had limited variation across all four types of climate; mostly, participants felt that faculty members in their department had a lot of autonomy with respect to teaching.

Table 2.6. Estimated Model Proportions and Estimated Means for the VEI model (*n*=166). The color coding in the table represents the average level of agreement on items within the construct, red indicating strong disagreement and green indicating strong agreement.

Constructs	Type of climate (Estimated model proportion)			
	Negative (13%)	Slightly positive (38%)	Positive (33%)	Very Positive (16%)
Involvement	2.4	3.7	4.0	4.5
Growth	1.9	2.9	3.3	4.1
Autonomy	4.2	3.7	4.1	4.1
Supervisor Support	2.6	3.3	4.1	4.6
Innovation	2.3	3.3	3.5	4.4
Outward Focus	2.2	3.3	3.9	4.4
Achievement	3.1	3.7	4.2	4.7
Performance feedback	1.7	2.4	3.3	4.1

As the color code provided in Table 2.6 indicates, we see an evolution between climates 1 to 4 from a negative to a positive description of the psychological collective climate around teaching. Thirteen percent of the participants fell into the first climate, which we label *Negative psychological collective climate around teaching*. Participants in this cluster indicated disagreeing with most of the constructs. In particular, they felt a lack of emphasis on personal development with respect to teaching (M_{Growth}=1.9) and a limited ability to get feedback on their teaching ($M_{Performance\ feedback}$=1.7). Over a third of the participants (38%) fell into the second climate, which we label *Slightly positive psychological collective climate around teaching*. Except for Performance Feedback, the mean across the constructs ranged from 2.9 to 3.7, indicating neutral to positive assessment of these constructs. A third of the participants (33%) belong to the third climate, which we label *Positive psychological collective climate around teaching,* with construct means ranging from 3.3 to 4.2. Finally, the last climate accounts for 16% of the participants. We label this fourth climate *Very positive psychological collective climate around teaching* since the construct means within this climate type are all above 4.0. Participants in this climate type reported that faculty members within their department were extremely involved with teaching-related decisions ($M_{Involvement}$=4.5), felt strongly supported by their chair ($M_{Supervisor\ support}$=4.6), and had a desire to excel in their teaching ($M_{Achievement}$=4.7). We explored through Fisher's exact tests whether academic rank and tenure status were related to the type of climate and found no statistically significant relationship for either (see Appendix C6 and C7).

While we were able to identify different types of psychological collective climate around teaching among our diverse population of faculty members, it is unclear whether a *departmental collective climate* around teaching exists within the departments surveyed. In the next section, we explore our third research question by leveraging data collected from the three departments that provided the highest response rate.

Departmental collective climate around teaching

Three departments coded as Department 16, Department 17, and Department 22 were chosen for this analysis since they had the highest response rates across our sample (90%, 77%, and 53% respectively before data cleaning; 71%, 46%, and 44% respectively after data cleaning; see Appendix C3 and C3). Although we were unable to have every single faculty member in these departments answer the survey, those that answered provide an adequate representation by academic rank (e.g., assistant, associate professor) of the department's composition (see Appendix C5).

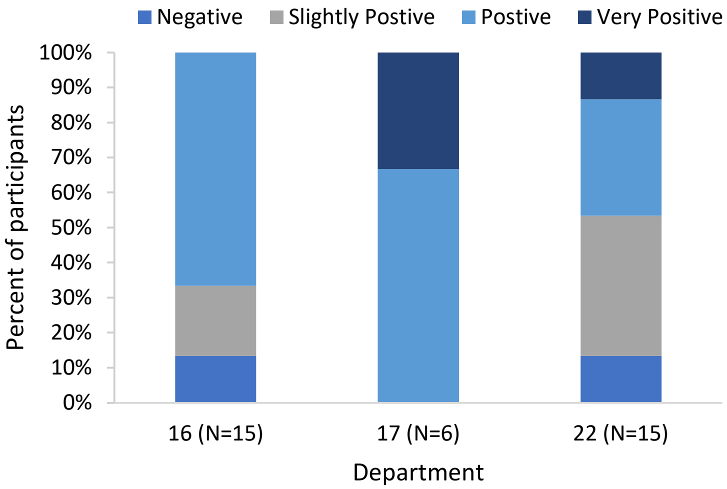


Figure 2.2 Distribution of types of psychological collective climate within high response rate departments.

In Figure 2.2, we provide a description of the variety of psychological collective climate present in each of these three departments. All three departments had a different distribution across the four types of psychological collective climate. All faculty members in Department 17 held a positive view towards the psychological collective climate around teaching in their department, while 13% of the faculty members in Department 22 and 16 had negative views about their departments.

As indicated in the theoretical framework, the psychological collective climate measures a faculty member's assessment of other faculty members' perceptions of the teaching climate within the department. This measure helps us understand the different points of view within a department but does not describe the climate around teaching of the department as a whole, i.e., the shared values. To obtain the latter, *departmental collective climate around teaching*, the level of consensus on the psychological collective climate among faculty members within the same department need to be tested. This is typically assessed in the literature by calculating the inter-rater agreement index $r_{wg(J)}$ and intraclass-correlation indices ICC(1) and ICC(2). $r_{wg(J)}$ indicates the level of agreement among faculty members within the same department with values below 0.75 indicating low agreement (O'Neill, 2017). As Table 2.7 indicates, although there was consensus within each department on most constructs, none of the departments had consensus on all constructs. Intraclass-correlation indices do not investigate absolute agreement but consistency between faculty members within a department. ICC(1) indicates whether the mean of one faculty is a reliable measure of the mean of another faculty within the same department. ICC(2) indicates whether there is a reliable difference between means across different departments. Results for both of these indices are presented in Table 2.8, along with associated cut-off interpretations. These results indicate that only Outward Focus met the inter-reliability criteria. Considering all three indices together, the only construct that could be aggregated was Outward Focus for this particular data set. Since only one of the eight constructs considered could provide the desired level of reliability, we conclude that departmental collective climate around teaching could not be measured for our three departments.

Table 2.7. $r_{wg(J)}$ results for high response rate departments. $r_{wg(J)}$ values below the 0.75 threshold are bolded.

Department	Involvement	Growth	Autonomy	Supervisor Support	Innovation	Outward Focus	Achievement	Performance feedback
16	**0.114**	0.779	0.886	0.810	0.845	0.771	0.790	**0.400**
17	0.872	0.805	0.876	0.928	0.775	0.954	0.949	**0.618**
22	**0.735**	0.853	0.938	0.897	0.868	0.871	0.928	**0.734**

Table 2.8. ICC values. ICC(1) value interpretation: >0.05 small to medium; >0.25 large effect; ICC(2) value interpretation: <0.4 poor; 0.40-0.75 fair to good; >0.75 excellent. Values that met the highest cut-off are bolded.

Construct	ICC(1)	ICC(2)
Involvement	0.067	0.410
Growth	0.139	0.610
Autonomy	-0.088	-3.595
Supervisor Support	0.017	0.140
Innovation	0.005	0.043
Outward Focus	**0.268**	**0.781**
Achievement	0.087	0.480
Performance feedback	-0.041	-0.611

Relationships between psychological collective climate, departmental collective climate and instructors' use of learner-centered instructional practices

The last research question focused on characterizing the relationship between the two different types of climate measure (i.e., psychological collective climate and departmental collective climate around teaching) and the level of use of learner-centered instructional practices. The abbreviated MIST instrument was used to measure the latter. Since we could not reliably measure departmental collective climate around teaching, we could not explore its relationship to instructional practices.

A simple linear regression was employed to predict instructors' use of learner-centered instructional practices based on the types of psychological collective climate (Eq. 2). Since some participants did not answer the abbreviated MIST portion of the survey, the number of participants eligible for this analysis was 149. A non-significant regression - F(3,145) = 1.029, p = 0.382 with an R^2 = 0.021 - indicated that in our data set, faculty members' view of psychological collective climate around teaching within their department could not be used to predict instructors' use of leaner-centered instructional practices. We went back to the cognitive interviews to identify whether faculty members provided some information that could point to a preliminary explanation for this lack of relationship. Several themes emerged that relate to a lack of communication about teaching among faculty members and lack of teaching-related standards. First, six of the eleven interviewees indicated that teaching was an independent endeavor in their departments:

> *"So, so inside the room I can, […] my students would be doing activities or whatever, but I know somebody else would just be, it'd be 50 minutes of lecture. So like we, and nobody would say anything about either of us. We were just doing our own thing."* Participants #3

> *"I think people tend to want to solve their own problems and figure things out for themselves. Everybody's fairly independent."* Participants #7

> *"I think the idea is giving people a lot of autonomy and I guess maybe we're, maybe we're sort of airing too far in the direction of autonomy and not enough on working together to think about best teaching practices, I think. I think it's just that it's kind of the culture in our department to give people as much autonomy as possible in terms of how they teach and what they choose to do."* Participants #7

The survey results also clearly demonstrated the high level of autonomy that faculty members have with respect to teaching (Table 2.6). Second, six of the eleven participants indicated that there were no expectations for someone to improve their teaching:

> *"I don't know that there is any sort of top-down expectations that that is going to be happening with any specific frequency. Um, or that you have to demonstrate on any sort of a regular basis that you have gotten better over some sort of times."* Participants #8

"I guess people are expected to get better, but like it doesn't matter if you don't either." Participant #1

Along the same vein, five indicated that there was no consensus or standard guidelines for the teaching approach one should employ:

"I don't think that as a department we make decisions about teaching methods." Participant #8

"Did we set high standards, you know, on some level, but like nobody could tell you what that standard is." Participant #3

"Like when I became an instructor, no one said, you know, you have to use a specific teaching method, right? Like, no one gave me any direction about that. […] I don't even think that the department, you know, has guidelines for how to teach." Participant #11

Discussion

This study provides insight into the challenges in measuring climate around teaching at the departmental level and highlights the need for a more rigorous approach to measuring this construct when exploring relationships between climate and other characteristics of the department such as the uptake of learner-centered instructional practices.

Cognitive interviews helped identify issues that need to be considered when measuring teaching climate

The cognitive interviews revealed challenges with constructs often measured in surveys on organization climate. For example, the construct Resources, which was included in two of the latest STEM-focused climate surveys (Landrum, et al., 2017; E. Walter, et al., 2014) had to be removed from the survey since interviewees indicated that resources (e.g., teaching and learning assistants, budgets) are typically not decided at the department level. Similarly, the construct of External Rewards was included in four of the twelve surveys listed in Table 1.1. The literature points to the need to include this construct as poor reward policies for teaching are often described as barriers to adoption of learner-centered instructional practices (Sturtevant & Wheeler, 2019). In particular, faculty members in these studies typically identify a lack of reward for teaching or a heavier emphasis on research during evaluation processes. The analysis of the cognitive interviews aligned with these findings. However, the interviewees also helped us realize that rewards and evaluation of teaching are not always decided and controlled at the department level, especially at four-year institutions. Moreover, the interviewees pointed to a lack of definition and tools to measure effective teaching. This is a common weakness of the higher education system in the United States and several initiatives are attempting to address it (Debad, 2020; National Science Foundation, 2020). Without rigorous and validated means to measure teaching, it is challenging to reward it. Therefore, the development of these tools is paramount for extrinsic rewards to be meaningfully included as a construct when measuring teaching climate at the institutional level and for extrinsic rewards to effectively impact the climate around teaching within a higher education institution.

Finally, the cognitive interviews highlighted the need to provide a timeframe for the survey participants to consider. The literature has highlighted that contrary to culture, climate can change over a shorter time scale (Schneider, et al., 2013). Moreover, studies have shown that academic leaders such as chairs and deans can influence climates within their unit (Kezar, 2016; Ramsden, Prosser, Trigwell et al., 2007). For example, a senior faculty with 10+ years in one department may have experienced different chairs and thus may respond to the climate survey based on their overall experience across these chairs or based only on the most recent chair. Junior faculty on the other end may only have experienced one chair. Therefore, providing a timeframe that is common to all survey participants would enhance the reliability and validity of the data. One caveat with focusing on just one chair is that senior faculty's assessment of the climate under that chair would be relative to other chairs they experienced in the past. It is thus critical to have broad representation of faculty across ranks when measuring the departmental climate.

It is important to define the level of the organization targeted when measuring climate

In this study, we leveraged the organization climate literature (e.g. studies in K-12 and industry) to investigate two different measures of climate based on the level of the organization that we were interested in: psychological collective climate for the individual level and departmental collective climate for the department. This contrasts with approaches in prior DBER studies, which measured psychological collective climate but treated it as a measure of departmental collective climate. We demonstrated that different types of psychological collective climate exist within a department, further reinforcing the assertion that this measure does not represent the departmental collective climate. Moreover, only one construct in our study met within-group consensus criteria required to measure departmental collective climate. These results thus demonstrate the challenges in measuring departmental collective climate in STEM departments.

In this study, we also leveraged the K-12 and industry literature on organizational climate to identify the set of constructs that defines organizational climate. However, the cognitive interviews indicated that some of these constructs (e.g., Resources and Extrinsic rewards) are not relevant for a climate measure at the department-level but would be meaningful to integrate in a climate measure at the college or institution level. Consequently, the relevancy of constructs to be included in a climate measure is dependent on the level of the organization targeted and should be explored during the initial stages of the development of a climate instrument.

Overall, future studies should carefully align the goals of their investigation and their measure of climate.

The link between climate around teaching within a department and faculty members' use of learner-centered instructional practices is more unclear than previously thought

One of the main goals of this study was to investigate the relationship between the uptake of learner-centered instructional practices and departmental climate around teaching measured at two different levels (i.e., the individual faculty and the department as a whole). We found a lack of relationship at the individual level. This is counter to numerous studies in which faculty members surveyed or interviewed had identified elements of departmental climate as barriers to instructional innovation (C. Henderson & M. H. Dancy, 2007; Landrum, et al., 2017; Sturtevant & Wheeler, 2019). These studies had led to the assumption that a departmental climate around teaching could inhibit or support the adoption of learner-centered instructional practices. In our study, we do not find evidence to support this assumption. Although exploring the reasons behind these findings was beyond the scope of this study, data collected here identify a potential reason: teaching is an autonomous endeavor with unclear expectations. First, we see in the survey responses and in the cognitive interviews that faculty members have a high level of autonomy when it comes to instructional strategies employed in their courses. These levels resonate with findings in previous studies (Landrum, et al., 2017; E. Walter, et al., 2014). Second, the interviews pointed to a lack of benchmark for effective teaching at the departmental level and the idea that instructional style is not imposed on faculty unless the evaluation process (e.g., student evaluation) identifies serious problems. Finally, the interviewees pointed to a lack of regular and frequent communication around teaching among departmental colleagues (as illustrated by the participants' inability to answer items related to informal conversations around teaching, Warmth construct). There is thus no shared understanding of effective teaching at the department level and limited opportunities to share best practices.

The definition of organizational climate, "the shared perceptions of and the meaning attached to the policies, practices, and procedures employees experience and the behaviors they observe getting rewarded and that are supported and expected," applied to teaching implies that there are policies around teaching and that effective teaching is rewarded. However, as we have seen in this study, effective teaching is ill-defined in most of the departments of our interviewees and consequently not measured; this results in teaching not being rewarded. Results in this study thus indicate that faculty within STEM departments have an underdeveloped departmental climate around teaching both at the individual and departmental level. However, future studies should delve more on this topic by addressing some of the limitations of the present study.

One implication of this finding is the necessity for higher education researchers to provide evidence-based descriptions as well as valid and reliable measures of effective teaching. Participants in this study echoed what has been reported in other studies, i.e., there is a lack of rewards and recognition of teaching. It may be that the department would reward teaching effectiveness if they felt they had the right tools to measure it. Having a set of criteria, albeit probably imperfect, for effective teaching and ways to measure it would also enable the development and value of high-profile teaching-related awards like those that exist for research (which are also based on an imperfect set of criteria). Moreover, if awards of the same stature as research awards existed for teaching, it may be that faculty members would consider them. In other words, the lack of focus on rewarding teaching may not be due to an unfavorable climate to rewarding teaching but rather the inexistence of high-profile ways to recognize it.

In turn, if departments and institutions leveraged this work to develop an evidence-based definition of effective teaching that aligns with their context, the conversations alone that would be required to achieve this vision would be a tremendous departure from what we have seen in this data set in terms of communication and involvement of faculty in teaching-related issues. It would help establish some of the elements of an organizational collective climate around teaching. Of course, if the adoption of such definition could lead to actionable policies that are understood, valued, and enforced by community members then we could be in a position where organizational collective climate can actually be measured.

Limitations

Several limitations should be considered for this study. First, our sample size, even if it met the criteria for conducting CFA analysis (at least five participants per item), is small (n=166). Moreover, the response rates at the department-level varied widely. The response rates of the three departments that were used to measure departmental collective climate were the highest in our sample but lower than what would be desired (>80%). These limitations associated with the sample size thus minimize the generalizability of the results. Second, the sample is biased toward research intensive institutions and chemistry departments. Eighteen of the twenty-two departments included in this study are embedded in research intensive institutions, nineteen are chemistry departments. The results should thus be viewed in that light. However, we included faculty members from primarily undergraduate institutions as part of the development and validation process of the DCaT survey and thus believe that the DCaT survey is well suited to be implemented in these environments. Third, our sampling strategy may have skewed the results toward more positive descriptions of climate since several of the faculty members within these departments had engaged in pedagogical professional development programs. This may have increased our ability to establish the validity of the data collected by the DCaT. Finally, we chose a Likert-type scale for this survey as we were interested in measuring unobservable characteristics of the faculty (i.e., their perceptions of their colleagues' perceptions of their departmental climate) and this option format is well-suited for large scale explorations of this type of characteristics (Ho, 2017). However, this type of scale is prone to biases such as central tendency bias, acquiescence bias, and social desirability bias (Subedi, 2016). Moreover, the 5-point scale provides low level of variations when compared to a quantitative scale.

Use of the DCaT survey

We have demonstrated that the DCaT survey can measure psychological collective climate but it is still unclear whether departmental collective climate is measurable with this instrument or other climate instruments. Caution should thus be taken when one implements this survey. If the intent is, for example, to monitor changes of the climate around teaching in a department as a reform effort is implemented, the DCaT survey can assist by evaluating changes in the distribution of types of psychological collective climate over time.

Evidence collected in this study were based on a small and bias sample. Implementation of the DCaT survey in a broader and larger sample would help further explore whether departmental collective climate is indeed not measurable. It would also be interesting to collect the DCaT survey along with a measure of climate around teaching at the College or institution level to explore overlaps between these two constructs and their differentiated abilities to explain uptake of learner-centered instructional practices. These studies as well as any studies employing the DCaT survey for research should conduct validity and reliability checks (especially cognitive interviews).

Whether the DCaT survey is used by a single department to understand their own climate around teaching or as part of a research project, we recommend triangulating the DCaT results with interviews in order to better understand the climate itself and factors influencing it. The interviews could explore in more depth some of the constructs measured in the DCaT survey; it could also probe other aspects of the climate not captured by the DCaT survey such as the role of extrinsic as well as intrinsic rewards regarding teaching; finally, the interviews could also shine some light on the influence of factors external to the department in shaping the climate around teaching (e.g., policies at the institutional and/or college level, and role of professional organization or accreditation agencies).

Conclusions

This study aimed to 1) measure departmental climate around teaching at the individual and department level using a newly developed instrument (DCaT) that has undergone rigorous validity and reliability studies, and 2) explore the relationship between these measures of departmental climate and uptake of evidence-based instructional practices within the department. Analyses of surveys collected from 166 faculty members representing twenty-two STEM departments at research intensive institutions show that 1) departmental climate around teaching is challenging to measure and clear operationalization of what is being measured is necessary, and 2) measure of departmental climate around teaching at the individual level (i.e., psychological collective climate) was not related to uptake of learner-centered instructional practices. This study is the first attempt at measuring departmental climate around teaching as defined in the organizational literature and testing a link between climate and instructional practices. Our findings suggest that departmental collective climate around teaching may be difficult to measure because most elements that define a climate (e.g., policies, practices, expectations) are lacking when it comes to teaching. Absence of these elements may contribute to the highly autonomous and independent approach to teaching that is seen in higher education and thus the lack of instructional innovation at scale.

List of abbreviations

CFA: Confirmatory Factor Analysis

CFI: Comparative Fit Index

CSC NFW: Cottrell Scholar Collaborative New Faculty Workshop

DBER: Discipline-Based Education Research

DCaT: Departmental Climate around Teaching

ICC: Intra-class correlation coefficients

MIST: Measurement Instrument for Scientific Teaching

MLR: Robust Maximum Likelihood

MMC: Mixture model clustering

POGIL: Process Oriented Guided Inquiry

RMSEA: Root Mean Square Error of Approximation

SRMR: Standardized Root Mean Square Residual

STEM: Science, technology, engineering, and mathematics

WIDER: Widening Implementation and Demonstration of Evidence Based Reforms

WLSMV: Weighted Least Square Mean and Variance adjusted

References

Ab Hamid, M. R.,Mustafa, Z.,Idris, F., et al. (2011). Measuring Value –Based Productivity: A confirmatory factor analytic (CFA) approach. *International Journal of Business and Social Science, 2*(6), 85-93.

American Educational Research Association,American Psychological Association,National Council on Measurement in Education, et al. (1999). *Standards for educational and psychological testing*: Amer Educational Research Assn.

Anderson, W. A.,Banerjee, U.,Drennan, C. L., et al. (2011). Changing the culture of science education at research universities. *Science, 331*(6014), 152-153. doi:10.1126/science.1198280

Arjoon, J. A.,Xu, X.,Lewis, J. E. (2013). Understanding the state of the art for measurement in chemistry education research: Examining the psychometric evidence. *Journal of Chemical Education, 90*(5), 536-545. doi:10.1021/ed3002013

Austin, A. E. (2011). *Promoting evidence-based change in undergraduate science education.* Paper presented at the Fourth committee meeting on status, contributions, and future directions of discipline-based education research.

Bathgate, M. E.,Aragon, O. R.,Cavanagh, A. J., et al. (2019). Supports: A key factor in faculty implementation of evidence-based teaching. *CBE Life Sci Educ, 18*(2), ar22. doi:10.1187/cbe.17-12-0272

Borda, E.,Schumacher, E.,Hanley, D., et al. (2020). Initial implementation of active learning strategies in large, lecture STEM courses: lessons learned from a multi-institutional, interdisciplinary STEM faculty development program. *International Journal of STEM Education, 7*(1). doi:10.1186/s40594-020-0203-2

Bouwma-Gearhart, J. (2012). Research university STEM faculty members' motivation to engage in teaching professional development: Building the choir through an appeal to extrinsic motivation and ego. *Journal of Science Education and Technology, 21*(5), 558-570.

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*: Guilford Publications.

Browne, M. W.,Cudeck, R. (1993). Alternative ways of assessing model fit. *Sociological Methods and Research, 21*, 230-258.

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, 11*(4), 339-346.

Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of applied psychology, 83*(2), 234.

Childs, P. E. (2009). Improving chemical education: turning research into effective practice. *Chemistry Education Research and Practice, 10*(3), 189-203.

Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of consulting and clinical psychology, 55*(4), 584.

Corbo, J. C.,Reinholz, D. L.,Dancy, M. H., et al. (2016). Framework for transforming departmental culture to support educational innovation. *Physical Review Physics Education Research, 12*(1), 010113.

Cottrell Scholars Collaborative. (2017). CSC New Faculty Workshop. Retrieved from http://chem.wayne.edu/feiggroup/CSCNFW/

Debad, S. J. (2020). Recognizing and evaluating science teaching in higher education: Proceedings of a workshop. *National Academies Press*.

Durham, M. F.,Knight, J. K.,Bremers, E. K., et al. (2018). Student, instructor, and observer agreement regarding frequencies of scientific teaching practices using the Measurement Instrument for Scientific Teaching-Observable (MISTO). *International Journal of STEM Education, 5*(1), 1-15.

Durham, M. F.,Knight, J. K.,Couch, B. A. (2017). Measurement Instrument for Scientific Teaching (MIST): A tool to measure the frequencies of research-based teaching practices in undergraduate science courses. *CBE—Life Sciences Education, 16*(4), ar67.

Eagan, K.,Stolzenberg, E. B.,Lozano, J. B., et al. (2014). Undergraduate teaching faculty: The 2013–2014 HERI faculty survey. . *Los Angeles: Higher Education Research Institute, UCLA*.

Ehrhart, M. G.,Schneider, B. (2016). Organizational climate and culture. In *Oxford Research Encyclopedia of Psychology*: Oxfor University Press.

Elrod, S.,Kezar, A. (2016). Increasing student success in STEM: A guide to systemic institutional change. *Washington, DC: AAC&U*.

Fraley, C.,Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Jornal, 41*(8), 578-588.

Froyd, J. (2011). *Propagation and realization of educational innovations in systems of undergraduate STEM education.* Paper presented at the Characterizing the Impact of Diffusion of Engineering Education Innovations Forum, National Academy of Engineering.

Furnham, A.,Goodstein, L. D. (1997). The organizational climate questionnaire (OCQ). *ANNUAL-SAN DIEGO-PFEIFFER AND COMPANY, 2*, 163-182.

Gess-Newsome, J.,Southerland, S. A.,Johnston, A., et al. (2003). Educational reform, personal practical theories, and dissatisfaction: The anatomy of change in college science teaching. *American Educational Research Journal, 40*(3), 731-767.

Hage, J.,Aiken, M. (1967). Relationship of centralization to other structural properties. . *Administrative Science Quarterly*, 72-92.

Halpin, A. W.,Croft, D. B. (1963). The organizational climate of schools *Administrator's Notebook, 11*(7).

Henderson, C.,Dancy, M. H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics-Physics Education Research, 3*(2), 020102.

Henderson, C.,Dancy, M. H. (2011). *Increasing the impact and diffusion of STEM education innovations.* Paper presented at the White paper commissioned for White paper commissioned for the Characterizing the Impact and Diffusion of Engineering Education Innovations Forum.

Ho, G. W. (2017). Examining perceptions and attitudes: A review of Likert-type scales versus Q-methodology. *Western journal of nursing research, 39*(5), 674-689.

Hora, M. T. (2012). Organizational factors and instructional decision-making: A cognitive perspective. *The Review of Higher Education, 35*(2), 207-235.

Hu, L. t.,Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

Kezar, A. (2016). Bottom-Up/Top-Down Leadership: Contradiction or Hidden Phenomenon. *The Journal of Higher Education, 83*(5), 725-760. doi:10.1080/00221546.2012.11777264

Knorek, J. K. (2012). Faculty teaching climate: Scale construction and initial validation. *PhD thesis, University of Illinois at Urbana-Champaign*.

Komperda, R.,Pentecost, T. C.,Barbera, J. (2018). Moving beyond Alpha: A Primer on Alternative Sources of Single-Administration Reliability Evidence for Quantitative Chemistry Education Research. *Journal of Chemical Education, 95*(9), 1477-1491. doi:10.1021/acs.jchemed.8b00220

Koo, T. K.,Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med, 15*(2), 155-163. doi:10.1016/j.jcm.2016.02.012

Landau, S.,Chis Ster, I. (2010). Cluster analysis: Overview. In *International Encyclopedia of Education* (pp. 72-83).

Landrum, R. E.,Viskupic, K.,Shadle, S. E., et al. (2017). Assessing the STEM landscape: the current instructional climate survey and the evidence-based instructional practices adoption scale. *Int J STEM Educ, 4*(1), 25. doi:10.1186/s40594-017-0092-1

Lund, T. J.,Stains, M. (2015). The importance of context: an exploration of factors influencing the adoption of student-centered teaching among chemistry, biology, and physics faculty. *International Journal of STEM Education, 2*(1), 1-21.

MacCallum, R. C.,Widaman, K. F.,Sehee Hong, S. Z. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84-99.

Mamiseishvili, K.,Lee, D. (2018). International faculty perceptions of departmental climate and workplace satisfaction. *Innovative Higher Education, 43*(5), 323-338. doi:10.1007/s10755-018-9432-4

National Science Foundation. (2013). Widening Implementation & Demonstration of Evidence-Based Reforms (WIDER). Retrieved from http://www.nsf.gov/pubs/2013/nsf13552/nsf13552.htm

National Science Foundation. (2020). Transforming higher education - Multidiemnsaionl evaluation of teaching. Retrieved from https://teval.net/index.html

Ngai, C.,Pilgrim, M. E.,Reinholz, D. L., et al. (2020). Developing the DELTA: Capturing cultural changes in undergraduate departments. *CBE Life Sci Educ, 19*(2), ar15. doi:10.1187/cbe.19-09-0180

O'Neill, T. A. (2017). An overview of interrater agreement on likert scales for researchers and practitioners. *Front Psychol, 8*, 777. doi:10.3389/fpsyg.2017.00777

Ostroff, C. (1993). The effects of climate and personal influences on individual behavior and attitudes in organizations. *Organizational behavior and human decision processes, 56*(1), 56-90.

Patterson, M. G.,West, M. A.,Shackleton, V. J., et al. (2005). Validating the organizational climate measure: links to managerial practices, productivity and innovation. *Journal of Organizational Behavior, 26*(4), 379-408. doi:10.1002/job.312

Peterson, C. H.,Peterson, N. A.,Powell, K. G. (2017). Cognitive interviewing for item development: Validity evidence based on content and response processes. *Measurement and Evaluation in Counseling and Development, 50*(4), 217-223. doi:10.1080/07481756.2017.1339564

POGIL Team. (2019). POGIL. Retrieved from https://pogil.org/

Ramsden, P.,Prosser, M.,Trigwell, K., et al. (2007). University teachers' experiences of academic leadership and their approaches to teaching. *Learning and Instruction, 17*(2), 140-155. doi:10.1016/j.learninstruc.2007.01.004

Reinholz, D. L.,Apkarian, N. (2018). Four frames for systemic change in STEM departments. *International Journal of STEM Education, 5*(1), 1-10.

Reinholz, D. L.,Corbo, J. C.,Dancy, M., et al. (2017). Departmental action teams: supporting faculty learning through departmental change. *Learning Communities Journal, 9*, 5-32.

Robinson, O. C. (2013). Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative Research in Psychology, 11*(1), 25-41. doi:10.1080/14780887.2013.801543

Schneider, B.,Ehrhart, M. G.,Macey, W. H. (2013). Organizational climate and culture. *Annual review of psychology, 64*, 361-388.

Shadle, S. E.,Marker, A.,Earl, B. (2017). Faculty drivers and barriers: laying the groundwork for undergraduate STEM education reform in academic departments. *International Journal of STEM Education, 4*(1), 1-13.

Sharma, S.,Mukherjee, S.,Kumar, A., et al. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research, 58*(7), 935-943.

Stains, M.,Harshman, J.,Barker, M. K., et al. (2018). Anatomy of STEM teaching in North American universities. *Science, 359*(6383), 1468-1470.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *J Pers Assess, 80*(1), 99-103. doi:10.1207/S15327752JPA8001_18

Sturtevant, H.,Wheeler, L. (2019). The STEM faculty instructional barriers and identity survey (FIBIS): Development and exploratory results. *International Journal of STEM Education, 6*(1), 1-22.

Subedi, B. P. (2016). Using Likert type data in social science research: Confusion, issues and challenges. *International journal of contemporary applied sciences, 3*(2), 36-49.

Thomas, K. W.,Tymon Jr, W. G. (2009). Work engagement profile. *Cpp, Mountain View, Ca*.

Walczyk, J. J.,Ramsey, L. L.,Zha, P. (2007). Obstacles to instructional innovation according to college science and mathematics faculty. *Journal of Research in Science Teaching, 44*(1), 85-106.

Walter, E.,Beach, A.,Henderson, C., et al. (2014). Describing instructional practice and climate: Two new instruments. *In Transforming Institutions: 21st Century Undergraduate STEM Education Conference, 24*.

Wieman, C.,Deslauriers, L.,Gilley, B. (2013). Use of research-based instructional strategies: How to avoid faculty quitting. *Physical Review Special Topics-Physics Education Research, 9*(2), 023102.

Wieman, C.,Perkins, K.,Gilbert, S. (2010). Transforming science education at large research universities: A case study in progressxs. *Change: The Magazine of Higher Learning, 42*(2), 6-14. doi:10.1080/00091380903563035

CHAPTER 3 Exploring the complementarity of measures of instructional practices

Introduction

The development and implementation of measures that provide reliable and valid evidence for the quality of teaching is critical for the improvement of the learning experiences provided of students enrolled in Science, Technology, Engineering, and Mathematics (STEM) courses. This evidence is essential for instructors to assess the effectiveness of their teaching, monitor changes as they work to improve students' learning outcomes in their courses, and for them to be objectively recognized during evaluation and promotion proceedings. Moreover, institutions could leverage these measures to inform the implementation of targeted support structures for instructors. Finally, these measures would help institutions demonstrate to accreditation agencies and other stakeholders the quality of the learning environments they provide to their students. Unfortunately, the measures currently employed at most institutions are inadequate. An analysis of the promotion and tenure policies of 51 research universities revealed that the most common measures of teaching effectiveness were student evaluations and peer classroom observations (Dennin, Schultz, Feig et al., 2017). Indeed, at most universities, the primary metric is student course evaluations (Henderson, Turpen, Dancy et al., 2014; Shao, Anderson, & Newsome, 2007) despite extensive evidence of their inappropriateness. For example, it is well established that student evaluations are influenced by various factors not related to instructional quality including the instructor's identities, student demographics, and subject area (e.g., Fan, et al., 2019; Heffernan, 2022; Hornstein, 2017). Peer classroom observations are also problematic since they are often conducted without guidelines or observation protocols that are aligned with effective teaching practices, and the peers observing may not be well versed in these practices. A survey of over 1,000 instructional staff across eight institutions supported by the Association of American Universities STEM Education Initiative indicated that only 13% of respondents described the quality of the evidence collected to measure effective teaching as high (Dennin, et al., 2017). There is thus an incredible need to develop measures and frameworks for the assessment of teaching that enable instructors' growth towards targeted effective instructional practices (Bradforth, Miller, Dichtel et al., 2015; Dennin, et al., 2017; Wieman, 2015).

Discipline-based education researchers have had a keen interest in measuring instructional practices to enhance teaching evaluation processes and to monitor changes resulting from the implementation of instructional reform efforts. Consequently, they have developed tools such as surveys, observation protocols, and rubrics that aim to provide valid and/or reliable characterizations of an instructor's teaching practices.

Surveys can be collected at scale with minimal time required for input and analysis while capturing a range of practices such as in-class behaviors, assessment practices, and out-of-class activities. Some prominent examples include the Teaching Practices Inventory (TPI; Smith, Vinson, Smith et al., 2014; Wieman & Gilbert, 2014), the Postsecondary Instructional Practices Survey (PIPS, Walter, Henderson, Beach et al.), and the Measurement Instrument for Scientific Teaching (MIST; Durham, et al., 2018; Durham, et al., 2017). While these surveys can show alignment with other measures of instructional practices (Durham, et al., 2018; Smith, et al., 2014), they can also be prone to validity threats. For example, Ebert-May et al. (2011) demonstrated the discrepancy between self-report surveys and classroom observations of instructional practices within the context of the evaluation of a pedagogical workshop.

Observation protocols have been considered a robust alternative to surveys since biases from the instructors themselves are removed. Most of the observation protocols can be grouped in two categories (M. Dancy, Hora, Ferrare et al., 2014): 1) holistic observation protocols which require the observers to make judgments at the end of the class session by answering survey-like questions or writing descriptive narratives based on the filed notes taken from the class session (e.g., reformed teaching observation protocol, RTOP; Sawada, et al., 2002); and 2) segmented observational protocols which require the observers to capture elements of the classroom instruction within a certain time frame (e.g., Classroom Observation Protocol for Undergraduate STEM, COPUS; Smith, Jones, Gilbert et al., 2013). The drawbacks of these observations protocols are that they solely focus on in-class practices and require extensive resources (observers) and time (for training and analysis).

Rubrics can address some of the weaknesses of the surveys and observation protocols by removing the instructors' biases while also capturing more holistically the experience provided to the students in a course rather than just in the classroom. An example of such a rubric is the Leaner-Centered Teaching Rubrics (LCTR; Blumberg, 2008). As described by Maryellen Weimer (2002), this set of rubrics assess instructors' use of learner-centered teaching. In a learner-centered teaching environment, the focus is on learning: "what the student is learning, how the student is learning, the conditions under which the student is learning, whether the student is retaining and applying the learning, and how current learning positions the student for future learning" (Weimer, 2002, p. xvi). In this environment, the instructor guides, facilitates, and designs the learning experiences and is no longer simply transmitting information. The five LCTR rubrics align with the five dimensions that Weimer advocates need to change to achieve learner-centered teaching: The *Function of Content* (i.e., students develop disciplinary skills along with in-depth conceptual understanding and understand the relevance of these acquired skills and knowledge), the *Role of the Instructor* (i.e., the instructor is a facilitator as opposed to a conveyor of knowledge), the *Responsibility for Learning* (i.e., the instructor fosters students' responsibility for learning), the *Purposes and Processes of Assessment* (i.e., assessments are on-going and promote reflection and learning), and the *Balance of Power* (i.e., the students have some control over the learning process). The LCTR rubrics were originally designed as a tool to help faculty in one-on-one consultations with a pedagogical expert to reflect on their instructional practices and identify areas for improvement (Blumberg, 2008, 2016). These consultations included analysis of classroom artifacts (e.g., syllabus, exams, lecture notes, etc.), which constitute authentic evidence of students' experiences in the course.

Teaching evaluation frameworks advocate for the triangulation of teaching data across different measures (Association of American Universities, 2019; Simonson, et al., 2021). For example, one popular framework among academic institutions and organizations is the Benchmarks for Teaching Effectiveness, developed by the Center for Teaching Excellence at the University of Kansas (Center for Teaching Excellence, 2021a). This framework aims to measure teaching by evaluating various facets of instruction and leveraging a variety of sources of evidence. A multidimensional rubric was developed as part of this framework to guide academic departments in their approach to teaching evaluation (Follmer Greenhoot, Ward, Bernstein et al., 2020). The rubric includes seven dimensions: Goals, content, and alignment; Teaching practices; Achievement of learning outcomes; Classroom climate; Reflection and iterative growth; Mentoring and advising; and Involvement in teaching service, scholarship, or community. The framework provides guidelines for the type of evidence that can be collected to support the evaluation of each of these dimensions(Center for Teaching Excellence, 2021a). For example, the following pieces of evidence are suggested to help assess the 'Teaching practices' dimension: syllabus, a sample of course materials, class observations supported by an observation protocol, dialogue with the instructor, and students' ratings and comments. The amount and breadth of evidence that should be collected and the need to identify tools to evaluate this evidence can be overwhelming for instructors, departments, and institutions and may contribute to limited uptake of these teaching evaluation frameworks. One strategy to mitigate this potential "overload" is to characterize the complementarity of existing instruments in order to assist with the educated selection of instruments.

In this study, we pursue this strategy and explore the relationship between COPUS and the LCTR rubrics. COPUS was chosen because it is an observation protocol that has been adopted extensively by both researchers (TRESTLE, 2017) and instructors (Arts & Sciences Support of Education Through Technology, 2022) due to its ease of use and objective output (i.e., capturing the behaviors of instructors and students occurring every two minutes). However, COPUS only captures the in-class learning experience and does not explicitly measure its quality. The set of LCTR rubrics, on the other hand, provides a holistic characterization of students' learning experiences in a course and align with research on effective practices to support learning (Blumberg, 2008). We chose this tool rather than a survey because teaching evaluation frameworks often request the collection of course artifacts (Center for Teaching Excellence, 2021b). The LCTR is one of the few tools that we are aware of that provide a systematic, evidence-based approach to analyzing these course artifacts. The LCTR rubrics provide comprehensive evidence for two of the dimensions on the Benchmarks for Teaching Effectiveness rubric ('Goals, content, and alignment' and 'Teaching practices'). However, the LCTR rubrics are time and resource intensive because there is a need for extensive analyst training, as well as to collect and analyze a large amount of data (course artifacts along with classroom observations). While it is reasonable to hypothesize that the way an instructor engages students in the classroom is a good predictor of the students' experience in the course overall, this hypothesis is yet to be fully explored. If the hypothesis holds, then it would be more effective to only use one instrument (presumably the one being least resource intensive) while retaining confidence in the ability to capture the instructional practices enacted in the course overall. This would leave resources to collect and analyze other sources of evidence such as student learning outcomes or reflection statements from instructors. We were thus interested in understanding the extent to which in-class behaviors (captured with COPUS) relate to the way an instructor implements their course (captured with the LCTR rubrics). If the output of the instrument that may be easier to implement and more appealing to its users (COPUS) aligns well with the output of the more holistic, yet resource-intensive instrument (LCTR), then there would be a benefit in using COPUS as a proxy measure for both dimensions of the Benchmarks for Teaching Effectiveness rubric (i.e., Goals, content, and alignment; Teaching practices). This would minimize the burden on data collection and analysis for the instructors (i.e., one measure can be used to collect evidence across two different dimensions of effective teaching). If a lack of alignment is found, this will provide evidence for the need to collect both types of evidence. The overall research question explored in this study is thus:

To what extent do in-class instructional behaviors, as measured using COPUS, relate to how an instructor approaches the teaching of their course, as determined using the LCTR rubrics?

Methods

Participants

The participants in this study were recruited from two different professional development workshops. Participants were recruited after enrolling in the workshops. The first set of participants (n = 15) are new chemistry assistant professors (within the first three years of their appointment) who attended the New Faculty Workshop (NFW; Stains, et al., 2015). The NFW participants taught a variety of chemistry courses ranging from general chemistry to upper-level graduate courses. The second set of participants (n = 13) are STEM instructors who attended a pedagogical workshop provided at one research-intensive institution. The STEM instructors represented different STEM departments (e.g., entomology, earth and atmospheric sciences, astronomy, etc.) and taught courses at either the undergraduate or graduate level. The two sets of participants bring the total sample size to 28 college instructors. Demographic information for each set of participants is listed in Table 3.1.

Table 3.1. Descriptive demographics for the participants

Demographic variables		**NFW Instructors, n**	**STEM Instructors, n**	**Total**
Gender*	Female	10	7	17
	Male	5	6	11
Course level taught	Undergraduate	7	10	17
	Graduate	8	3	11
Teaching experience	Less than 3 years	13	1	14
	At least 3 years	2	12	14
Total		15	13	28

Participants self-identified on a survey that asked them to provide their demographic information

The study was approved by the University of Nebraska-Lincoln and the University of Virginia Institutional Review Boards.

Data collection

Participants' classroom video observations and course artifacts were collected during the semester following their participation in the workshop. Observations and course artifacts were associated with a particular unit/topic/chapter taught by each participant. Each participant was allowed to select the unit/topic/chapter to be analyzed for this study. The number of classroom video observations ranged from 2-8 per participant with a mode of 4. The course artifacts included the syllabus, any course material used to teach the selected unit/topic/chapter (e.g., slides, class notes, etc.), and any assessment tools used to assess the selected unit/topic/chapter (e.g., homework, quiz, mid-term/final exam, etc.).

Analysis of in-class learning experience using COPUS

The classroom videos were analyzed with the COPUS (Smith, et al., 2013). This protocol provides information about how instructors and students spend their time in the classroom. It requires observers to code every two minutes of class time for 13 students' behaviors (e.g., listening, answering questions, etc.) and 12 instructors' behaviors (e.g., lecturing, posing questions, etc.). COPUS does not require observers to make judgments of teaching quality, only recording the frequency of particular behaviors.

The researchers who used COPUS to code classroom videos in this study were trained as a cohort, with the training process led by other researchers who had conducted prior studies utilizing COPUS. The training and coding processes are summarized in the following steps:

Three sets of training videos were selected based on the degree of difficulty in coding them in prior coding sessions ("easy", "moderate", and "hard"). For the easy videos, most of the classroom time consists of the instructor lecturing and students listening. The hard videos include more varied forms of interactions between instructors and students, necessitating the use of a wide spectrum of COPUS codes.

During the training process, the researchers watched and coded the easy videos independently. Each researcher entered the results of their coding into a joint spreadsheet. This spreadsheet was used to facilitate the discussions of coding disagreements with the entire cohort of coders (N = 7). The coding spreadsheet was used to calculate Fleiss' Kappa, which allows for the determination of interrater agreement between more than two raters (Nichols, Wisner, Cripe et al., 2010). Fleiss' Kappa was calculated using the "irr" package in R (Gamer, Lemon, Gamer et al., 2012). If Fleiss' Kappa values for a classroom observation reached a threshold of above 0.8, the group shifted to training on the next harder set of videos. However, if the value was below 0.8, the process was repeated on a video of similar difficulty following the discussion of coding disagreements.

After reaching the desired level of agreement for all training videos, 106 classroom observations collected from the 28 participants were distributed across the seven trained coders. Each coded 10 to 20 videos.

Once all videos were coded, the spreadsheet containing the coding was submitted to the COPUS Analyzer (Stains, et al., 2018). This tool helps classify each observation into three instructional styles: didactic (i.e., the instructor lectures for, on average, more than 80% of class time), interactive lecture (i.e., the lecture is supplemented with some student-centered strategies such as group work, asking clicker questions), and student-centered (i.e., instructors incorporate student-centered strategies into a large portion of their class time for such that, on average, only about 50% of class time is spent lecturing). The COPUS Analyzer provided a classification into the three instructional styles for each video.

Since we were interested in describing the instructional style of the instructor overall and each instructor provided more than one video classroom observation, we represented the overall instructional style of an instructor by calculating the weighted average of the instructional styles observed across all the video classroom observations obtained from the instructor:

Weighted average COPUS for one instructor = (1 × percent of videos collected from the instructor classified as didactic) + (2 × percent of videos collected from the instructor classified as interactive lecture) + (3 × percent of videos collected from the instructor classified as student-centered) ***Eq.3.1***

The weighted average COPUS has a range from 1 to 3, with 1 representing instructors for whom all videos were classified as Didactic, and 3 representing instructors for whom all videos were classified as Student-centered.

Analysis of in- and out-of-class learning experiences using LCTR

The Learner-Centered Teaching Rubrics (LCTR; Blumberg, 2008, 2016) were utilized to more broadly characterize how an instructor taught a course. Course artifacts and two classroom observations for each participating instructor were used to assign scores on the rubrics. The five rubrics (the *Function of Content*, the *Role of the Instructor*, the *Responsibility for Learning*, the *Purposes and Processes of Student Assessment*, and the *Balance of Power*) each contain several components. Each component is measured on a four-point scale with 1 representing “Instructor-centered approach”, 2 and 3 representing “Lower” and “Higher level of transitioning” respectively, and 4 representing “Learner-centered approach”. The original rubrics (Blumberg, 2008) required some modifications for this study since some of the components were not measurable with the obtained classroom observations and course artifacts (Popova, et al., 2020). One example is that based on classroom observations collected and the course artifacts, which did not contain student answers, it was not possible to identify whether the students had opportunities to justify their answers when they did not agree with those of the instructor. This eliminated one component within the rubric *Purposes and Processes of Assessment.* The modified LCTR rubrics employed in this study contained 14 components across the five rubrics (Appendix D).

Three members of the research team were involved in the coding process. First, the researchers coded the selected classroom observations and course artifacts independently for 3-4 instructors. Then, the coders discussed their assigned scores for each component of each LCTR rubric to resolve any disagreements. After reaching a 100% agreement, the researchers coded the rest of the instructors independently. However, to ensure adherence to the rubrics over time, two researchers coded the same instructor after they had coded three or four other instructors independently. In total, 10 instructors were coded by two researchers and 18 by a single researcher.

For each instructor, the scores on the components within a rubric were averaged to obtain a rubric-level score. Then, the rubric-level scores were averaged to obtain, what we will refer to as the LCTR score. The LCTR scores ranged from 1 to 4 with 1 representing instructors using "Instructor-centered approach", 4 representing instructors using "Learner-centered approach", and 2 and 3 identified as "Lower level of transitioning" and "Higher level of transitioning" respectively.

Comparison of COPUS and LCTR classifications

In order to make the weighted average COPUS and LCTR scores comparable, a normalized weighted average COPUS and normalized LCTR score were calculated by using the following equations:

Normalized weighted average COPUS = (Weighted average COPUS -1)/ (3-1) **Eq.3.2**

Normalized LCTR score = (LCTR score -1)/ (4-1) **Eq.3.3**

Both normalized weighted average COPUS (NWA-COPUS) and normalized LCTR score (N-LCTR) have a range from zero to one, which enables facile comparisons between the output from the two tools. This range was divided into four categories for both NWA-COPUS and N-LCTR. The range of 0.00-0.25 is described as *Instructor-centered* for LCTR and *Didactic* for COPUS in order to match the language of previous studies (Blumberg, 2008; Stains, et al., 2018), but both classifications represent instructors using most of their class time to lecture; 0.26-0.50 refers to *Lower Interactive Lecture* for COPUS and *Lower Level of Transitioning* for LCTR; 0.51-0.75 refers to *Higher Interactive Lecture* for COPUS and *Higher Level of Transitioning* for LCTR; finally, the range 0.76-1.00 represents *Student-centered* for COPUS and *Learner-*

centered for LCTR. Detailed descriptions of each of these categories are provided in Appendix E and Appendix F.

Results

We share below the distribution of instructional styles across the 28 instructors as described by COPUS and the LCTR rubrics. We then provide the results of the alignment in the classification of instructional styles between COPUS and the LCTR rubrics. Finally, we report on the influence of observation intensity (i.e., the number of classroom observations conducted per participant) on this alignment.

Participants' instructional styles according to COPUS and the LCTR rubrics

As Figure 3.1a indicates, most of the instructors were classified as either *Didactic* (36%) or *Lower Interactive Lecture* (43%) based on the NWA-COPUS data. Only 7% were classified as *Student-centered*. The N-LCTR data provided a different distribution with two thirds of the instructors classified as *Lower Level of Transitioning*, a fifth as *Higher Level of Transitioning*, and none as *Learner-centered*.

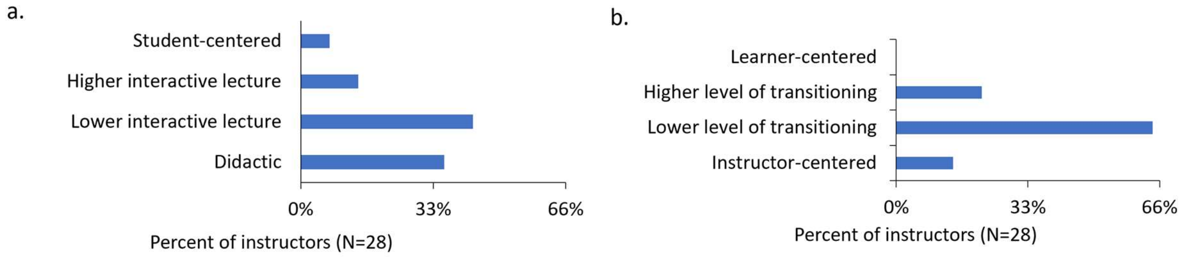


Figure 3.1 Classification of participants' instructional styles according to a) COPUS and b) the LCTR rubrics

Alignment in instructional styles between COPUS and the LCTR rubrics

The alignment between in-class learning experience (measured with COPUS) and in/out of class learning experiences (measured with the LTCR rubrics) in a course was probed further by grouping instructors by their COPUS instructional styles and then classifying each instructor within a COPUS style by their LCTR instructional style (Figure 3.2). Analysis of Figure 3.2 provides insight into the partial misalignment between the two measures observed in Figure 3.1. For example, in theory, we would expect to see a large percentage of the participants in the *Didactic* instructional style (COPUS) to be classified as *Instructor-centered* (LCTR). However, of the ten participants in the *Didactic* style, only four were classified as *Instructor-centered* by the LCTR rubrics. Similarly, the twelve participants classified as *Lower level of transitioning* with the LCTR rubrics were spread across three COPUS instructional styles (33% *Didactic*, 50% *Lower interactive*, and 17% *Higher interactive*). As an illustration of how the two instruments may classify the same instructor differently, we present a detailed description of the results of analysis for two individual instructors.

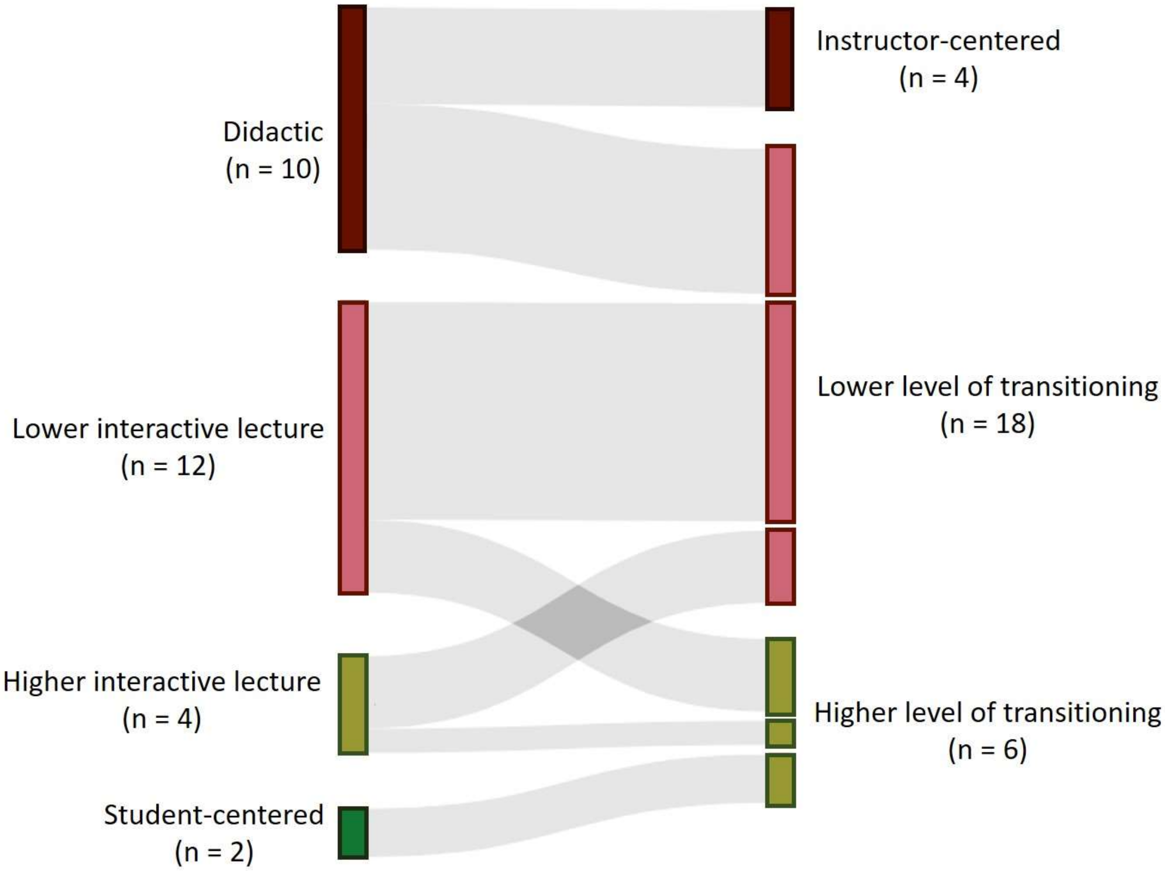


Figure 3.2 A Sankey diagram visualizing the alignment between the NWA-COPUS and N-LCTR classifications. The width of the nodes and edges provides quantitative flow information.

Instructor 1 was teaching an undergraduate engineering course and had been teaching for two years at the time of data collection. The COPUS analysis of the three videos that were collected indicated that they spent on average 93.1 ± 6.5% of class time lecturing. No group work was observed in these three videos. Based on these data, their instructional style was classified as *Didactic*. Analysis of course artifacts and videos with the LCTR rubrics led Instructor 1 to be classified as *Lower level of transitioning*, indicating that they were not classified solely as *Instructor-centered* across all five LCTR rubrics. For example, in the *Role of Instructor* rubric, they were classified as *Higher level of transitioning*. Indeed, they facilitated learning by using a variety of instructional strategies including minute papers (captured on video and described in the syllabus), group work (evidence from syllabus), and opportunities for students to critique each other's work (evidence from syllabus). They also had low- and high-level learning objectives listed on the syllabus and were observed talking to students about some of these learning objectives during class.

Instructor 2 was teaching an undergraduate astronomy course and had been teaching for three years at the time of data collection. The LCTR analysis of their course artifacts and videos classified them in *Lower level of transitioning*. However, this classification did not capture the nuances in the instructor's practices. For example, within the *Purposes and Process of Assessment* rubric, the instructor employed formative assessment in the form of Peer Instruction (a *Learner-centered* practice), but they did not use peer assessment or authentic assessment (an *Instructor-centered* practice), and ultimately were classified as *Lower level of transitioning* for this rubric. While the instructor obtained a *Lower level of transitioning* score on each of the five LCTR rubrics, the COPUS analysis of their four videos classified them as *Higher interactive*. They lectured for an average of 79.7 ± 15.4% of class time and group work took place in every session observed with an average of 17.8 ± 16.4% of class time. During one observation, they lectured for only 58% of the class and students worked in groups on a worksheet for 42% of the class time.

While Figure 3.2 displays a large number of misalignments, those are typically occurring within the same approximate instructional style. For example, although some participants classified as *Didactic* were placed in the *Lower level of transitioning* instructional style based on the LCTR rubrics, no one shifted from *Didactic* to *Higher level transitioning* or to *Learner-centered*.

The complete alignment observed between the two measures only occurred at the extremes of the scales. Four instructors who were classified as *Instructor-centered* with the LCTR rubrics were classified as *Didactic* based on the COPUS analysis of their classroom observations. Two instructors who were classified as *Student-centered* with COPUS were classified as *Higher level of transitioning* with the LCTR rubrics.

Upon further analysis of the alignment between the COPUS data and the LCTR rubrics, we noticed differences at the level of the LCTR rubrics. We found that the *Didactic* and *Lower interactive lecture* (COPUS classification) aligned well with scores on the LCTR rubric *Purposes and processes of assessment* (i.e., assessment should be given during the learning process) as Appendix C12 indicates. A *Didactic* classification also aligned well with scores on the *Responsibility for learning* (i.e., instructors can teach students to take responsibility for their learning) with 80% of the Didactic instructors being classified as *Instructor-centered* on this rubric (Appendix C11). However, some misalignments were observed. For example, three quarters of the participants who were classified as *Higher interactive lecture* (COPUS classification) were typically classified in comparatively lower levels on the following LCTR rubrics: *Responsibility for learning*, *Purposes and processes of assessment*, and *Balance of power*. On the other end, COPUS classifications underestimated scores on *The function of content* and *The role of the instructor* rubrics since at least 50% of the instructors in each COPUS classification were assigned to a comparatively higher level in the LCTR classification (Appendix C9 and Appendix C10).

Overall, the data show partial alignment between the two measures. Complete alignments were minimal and occurred at the extremes of the teacher-centered/learner-centered instruction spectrum. Misalignments were common but did not lead to contradicting classifications of instructional styles between the two measures. Instead, they captured nuances within broad instructional styles (i.e., didactic, intermediate, and student/learner centered). These results indicate that both instruments capture different aspects of teaching and that one could not be used as a direct substitute for the other.

Comparing NWA-COPUS and N-LCTR by observation intensity

Several studies, including ours, have remarked that the level of accuracy in the description of an instructor's in-class practice relies on coding several class sessions, although the minimum number of sessions to be observed and their timing is still being determined (Denaro, Sato, Harlow et al., 2021; Lund, Pilarz, Velasco et al., 2015; McConnell, Boyer, Montplaisir et al., 2021; Sbeglia, Goodridge, Gordon et al., 2021; Stains, et al., 2018; Weston, Hayward, & Laursen, 2021). We thus intended to explore whether the level of alignment between the LCTR and COPUS scores was related to the number of classroom observations coded with COPUS. Across the 28 instructors, the number of classroom observations ranged from 2-8 per instructor with a mode of 4 (Appendix C14). Participants were classified into five groups based on the number of observations collected (Appendix C14).

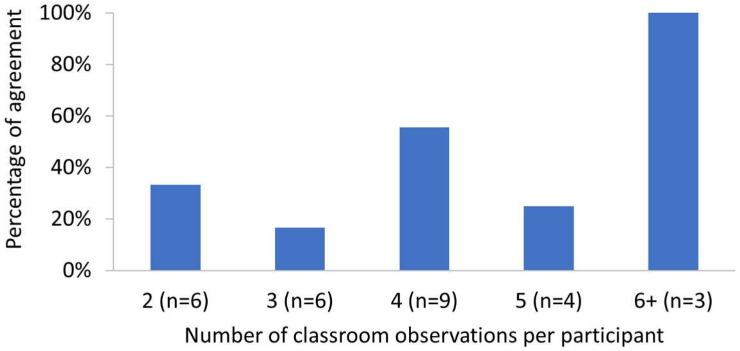


Figure 3.3 Proportion of instructors for which alignment existed between the N-LCTR and NWA-COPUS classifications by observation intensity.

For each of these five groups, we tabulated the proportion of instructors for which alignment existed between the N-LCTR and NWA-COPUS classifications. Figure 3.3 shows that perfect alignment was observed among the instructors (n = 3) with more than six classroom observations. These participants were classified in the two lower levels of each classification (*Didactic*/*Instructor-centered* or *Lower interactive lecture*/*Lower level of transitioning*). However, for the groups with lower observation intensity, the agreement between NWA-COPUS and N-LCTR ranged from 17% to 56% with no clear trend observed by sample intensity. However, all misalignments consisted of shifts of one level of classification above or below the level identified by one measure.

Discussion and Implications

Teaching evaluation frameworks have advocated for the use of multiple sources of evidence to characterize instructional practices (Association of American Universities, 2019; Center for Teaching Excellence, 2021a; Simonson, et al., 2021). Thanks to extensive efforts from the education research community, we now have a plethora of instruments that measure various aspects of teaching. Each comes with its strengths and weaknesses concerning the validity and reliability of the data collected as well as with the level of resources required to collect and analyze the data. To assist instructors, researchers, and institutions in their selection of instruments, it is important to explore and report on the complementarity, or lack thereof, between instruments. In this study, we sought to identify the complementarity between a popular observation protocol focused on instructors' and students' behaviors in the classroom, COPUS, and the LCTR rubrics, which measure the level of learner-centeredness of a course.

Our results show that the outcomes from COPUS and the LCTR rubrics exhibit partial misalignment, indicating that one instrument should not be used in place of the other. Rather each should be selected to align with the teaching dimension that is intended to be captured and claims should be limited to that dimension. For example, researchers interested in capturing the impact of a new professional development initiative focused on in-class group work would benefit from using COPUS and not LCTR. Similarly, if one of the main criteria in evaluating teaching effectiveness for an institution is the implementation of group work in the classroom, then COPUS is a better measure. Interestingly, the misalignments observed were moderate (i.e., an instructor was not classified as instructor-centered with one measure and student-centered with another measure), indicating that both measures agree on the course description of a faculty member's instructional style. Overall, the results of this study indicate that one should be cautious when relying on COPUS data to evaluate the extent to which an instructor implements learner-centered practices in their courses and vice versa.

While the results and their implications relate to the use of these two instruments for the purpose of summative evaluation, these instruments can also serve a formative purpose. Indeed, formative evaluation can be used to promote growth among instructors and each of these tools could be valuable depending on the nature of the growth expected. COPUS can be a great entry point when working with faculty with limited exposure and understanding of learner-centered teaching. It could be used, for example, in the formative years of new professors to help them become comfortable and develop an appreciation for the benefits of collaborative learning. For more seasoned instructors, the LCTR rubrics can help them reflect on and provide them with benchmarks to ensure a holistic, learner-centered approach to their courses.

Importantly, this study displays the complexity of measuring instructional effectiveness and supports the recommendation by teaching evaluation frameworks to employ several sources of evidence and analytical tools to describe an instructor's teaching practices (Association of American Universities, 2019; Center for Teaching Excellence, 2021a; Simonson, et al., 2021). Yet it is still unclear how to combine the multiple sources of evidence to provide an overall description and assessment of teaching quality for an instructor. In this study, we saw variations in the level of implementation of student/learner-centered practices across the two measures and within measures. For example, Instructors 1 and 2 scored high on certain components of an LCTR rubric but not on others. Moreover, Instructor 2 showed great variability in their use of group work in their classroom with one session having only 8% of class time spent on group work versus 42% in another session. It is quite challenging to the effectiveness of these two instructors even with the evidence collected here, which are quite extensive and robust compared to typical measures of teaching evaluations (Dennin, et al., 2017). Notably absent from the evidence collected for this study are measures of student learning. While the addition of this type of evidence could bring some clarity, providing too much weight to them could have its own pitfalls, especially if the measures are inappropriate (such as student evaluations). The development of rubrics such as those developed as part of the Benchmarks for Teaching Effectiveness framework (Center for Teaching Excellence, 2021a) addresses some of these challenges. However, more research is needed to explore how instructors, committee members, and administrators actually apply these rubrics.

In summary, this exploratory study highlights the need for institutions and education researchers to carefully consider the alignment between the tools they employ to characterize instructional practices and the nature of the practices desired. Misalignments can lead to an invalid assessment, which, in turn, can decrease the credibility of the evaluation process. It also demonstrates the challenges and messiness of measuring instructional quality.

Limitations

The small sample size limits the generalizability of the findings. At the same time, the diversity of instructors and courses represented alleviates some of the generalizability concerns. This study and associated findings should be considered exploratory and more extensive studies need to be conducted to reach generalizable results.

Conclusion

This research compared two instruments, Classroom Observation Protocol for Undergraduate STEM (COPUS) and the Learner-Centered Teaching Rubrics (LCTR), in measuring teaching practices including 28 STEM instructors from research-intensive institutions across the United States. The results indicated a partial misalignment in measuring teaching practices by utilizing the two instruments which highlighted the complexity of evaluating instructional practices and emphasize the need to choose appropriate evaluation tools for different purposes to ensure the credibility of the evaluation. Comparing the two instruments under different groups with different numbers of observations found no clear trend between the alignment of the two instruments and the sample intensity, and more research is needed to discuss sample intensity requirements for evaluating instructional practices.

References

Arts & Sciences Support of Education Through Technology. (2022). COPUS Tool Details. Retrieved from https://www.colorado.edu/assett/programs/vips/copus

Association of American Universities. (2019). AAU Undergraduate STEM Education Initiative: Matrix of summative evaluation of teaching strategies. Retrieved from https://www.aau.edu/sites/default/files/AAU-Files/STEM-Education-Initiative/P&T-Matrix.pdf

Blumberg, P. (2008). *Developing learner-centered teaching: A practical guide for faculty*: John Wiley & Sons.

Blumberg, P. (2016). Assessing Implementation of Learner-Centered Teaching While Providing Faculty Development. *College Teaching, 64*(4), 194-203. doi:10.1080/87567555.2016.1200528

Bradforth, S. E.,Miller, E. R.,Dichtel, W. R., et al. (2015). University learning: Improve undergraduate science education. *Nature, 523*(7560), 282-284.

Center for Teaching Excellence. (2021a). Benchmarks for teaching effectiveness. Retrieved from https://cte.ku.edu/benchmarks-teaching-effectiveness-project

Center for Teaching Excellence. (2021b). KU Benchmarks Rubric: Evidence Matrix, by Teaching Dimension and Source In: The University of Kansas.

Dancy, M.,Hora, M.,Ferrare, J., et al. (2014). *Describing and measuring undergraduate STEM teaching practice: A report from a national meeting on the measurement of undergraduate science, technology, engineering and mathematics (STEM) teaching.* Paper presented at the American Association for the Advancement of Science.

Denaro, K.,Sato, B.,Harlow, A., et al. (2021). Comparison of cluster analysis methodologies for characterization of Classroom Observation Protocol for Undergraduate STEM (COPUS) data. *CBE—Life Sciences Education, 20*(1), ar3.

Dennin, M.,Schultz, Z. D.,Feig, A., et al. (2017). Aligning practice to policies: Changing the culture to recognize and reward teaching at research universities. *CBE—Life Sciences Education, 16*(4), es5.

Durham, M. F.,Knight, J. K.,Bremers, E. K., et al. (2018). Student, instructor, and observer agreement regarding frequencies of scientific teaching practices using the Measurement Instrument for Scientific Teaching-Observable (MISTO). *International Journal of STEM Education, 5*(1), 1-15.

Durham, M. F.,Knight, J. K.,Couch, B. A. (2017). Measurement Instrument for Scientific Teaching (MIST): A tool to measure the frequencies of research-based teaching practices in undergraduate science courses. *CBE—Life Sciences Education, 16*(4), ar67.

Ebert-May, D.,Derting, T. L.,Hodder, J., et al. (2011). What We Say Is Not What We Do: Effective Evaluation of Faculty Professional Development Programs. *BioScience, 61*(7), 550-558. doi:10.1525/bio.2011.61.7.9

Fan, Y.,Shepherd, L. J.,Slavich, E., et al. (2019). Gender and cultural bias in student evaluations: Why representation matters. *PloS one, 14*(2), e0209749.

Follmer Greenhoot, A.,Ward, D.,Bernstein, D., et al. (2020). Benchmarks for Teaching Effectiveness. Retrieved from https://cte.ku.edu/sites/cte.ku.edu/files/docs/KU%20Benchmarks%20Framework%202020update.pdf

Gamer, M.,Lemon, J.,Gamer, M. M., et al. (2012). Package 'irr'. *Various coefficients of interrater reliability and agreement, 22*.

Heffernan, T. (2022). Sexism, racism, prejudice, and bias: a literature review and synthesis of research surrounding student evaluations of courses and teaching. *Assessment & Evaluation in Higher Education, 47*(1), 144-154.

Henderson, C.,Turpen, C.,Dancy, M., et al. (2014). Assessment of teaching effectiveness: Lack of alignment between instructors, institutions, and research recommendations. *Physical Review Special Topics - Physics Education Research, 10*(1). doi:10.1103/PhysRevSTPER.10.010106

Hornstein, H. A. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education, 4*(1), 1304016.

Lund,Pilarz,Velasco, et al. (2015). The best of both worlds: Building on the COPUS and RTOP observation protocols to easily and reliably measure various levels of reformed instructional practice. *CBE Life Sci Educ, 14*(2). doi:10.1187/cbe.14-10-0168

McConnell, M.,Boyer, J.,Montplaisir, L. M., et al. (2021). Interpret with caution: COPUS instructional styles may not differ in terms of practices that support student learning. *CBE—Life Sciences Education, 20*(2), ar26.

Nichols, T. R.,Wisner, P. M.,Cripe, G., et al. (2010). Putting the Kappa Statistic to Use. *The Quality Assurance Journal, 13*(3-4), 57-61. doi:10.1002/qaj.481

Popova, M.,Shi, L.,Harshman, J., et al. (2020). Untangling a complex relationship: Teaching beliefs and instructional practices of assistant chemistry faculty at research-intensive institutions. *Chemistry Education Research and Practice, 21*(2), 513-527.

Sawada, D.,Piburn, M. D.,Judson, E., et al. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School science and mathematics, 102*(6), 245-253.

Sbeglia, G. C.,Goodridge, J. A.,Gordon, L. H., et al. (2021). Are Faculty Changing? How Reform Frameworks, Sampling Intensities, and Instrument Measures Impact Inferences about Student-Centered Teaching Practices. *CBE—Life Sciences Education, 20*(3), ar39.

Shao, L. P.,Anderson, L. P.,Newsome, M. (2007). Evaluating teaching effectiveness: where we are and where we should be. *Assessment & Evaluation in Higher Education, 32*(3), 355-371. doi:10.1080/02602930600801886

Simonson, S. R.,Earl, B.,Frary, M. (2021). Establishing a framework for assessing teaching effectiveness. *College Teaching*, 1-18.

Smith, M. K.,Jones, F. H.,Gilbert, S. L., et al. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. *CBE Life Sci Educ, 12*(4), 618-627. doi:10.1187/cbe.13-08-0154

Smith, M. K.,Vinson, E. L.,Smith, J. A., et al. (2014). A campus-wide study of STEM courses: new perspectives on teaching practices and perceptions. *CBE Life Sci Educ, 13*(4), 624-635. doi:10.1187/cbe.14-06-0108

Stains, M.,Harshman, J.,Barker, M. K., et al. (2018). Anatomy of STEM teaching in North American universities. *Science, 359*(6383), 1468-1470.

Stains, M.,Pilarz, M.,Chakraverty, D. (2015). Short and long-term impacts of the Cottrell scholars collaborative new faculty workshop. *Journal of Chemical Education, 92*(9), 1466-1476.

TRESTLE. (2017). COPUS Observations Resources. Retrieved from https://trestlenetwork.ku.edu/copus-observation-resources/

Walter,Henderson,Beach, et al. Introducing the Postsecondary Instructional Practices Survey (PIPS): A Concise, Interdisciplinary, and Easy-to-Score Survey. *CBE Life Sci Educ, 15*(4). doi:10.1187/cbe.15-09-0193

Weimer, M. (2002). *Learner-centered teaching: Five key changes to practice*: John Wiley & Sons.

Weston, T. J.,Hayward, C. N.,Laursen, S. L. (2021). When seeing is believing: Generalizability and decision studies for observational data in evaluation and research on teaching. *American Journal of Evaluation, 42*(3), 377-398.

Wieman, C. (2015). A better way to evaluate undergraduate teaching. *Change: The Magazine of Higher Learning, 47*(1), 6-15.

Wieman, C.,Gilbert, S. (2014). The teaching practices inventory: a new tool for characterizing college and university teaching in mathematics and science. *CBE Life Sci Educ, 13*(3), 552-569. doi:10.1187/cbe.14-02-0023

CHAPTER 4: Explorations of General Chemistry Instructors' Rationales behind their Assessment Practices

Introduction

Student learning is mediated by different factors such as their motivation to learn (Zimmerman, Bandura, & Martinez-Pons, 1992), conceptions and approaches to learn (Prosser & Trigwell, 1999), perceptions of the learning environment (Gijbels, Segers, & Struyf, 2008; Trigwell & Prosser, 1991), and most importantly, the assessment of learning (Momsen, Offerdahl, Kryjevskaia et al., 2013; Wiliam, 2011). Assessment drives student learning because students selectively study the content and skills that will help them successfully pass the course (Hanauer & Bauerle, 2012; Momsen, et al., 2013). To support the propagation of new assessment approaches that can be used to assess students' conceptual understanding and scientific practices proposed by the chemistry education research community (Bretz, 2014; Stowe & Cooper, 2019), first, we need to understand instructors' rationales for assessing students and their assessment practices.

Purpose of assessment

Assessment served various purposes in evaluating or promoting students' learning. Earl (2006) grounded in literature to define the purpose of assessment from three aspects: assessment *for* learning, assessment *as* learning, and assessment *of* learning. Assessment *for* learning refers to instructors utilizing the information collected through the learning process to determine students' progress, provide feedback to students, and adapt their teaching practices accordingly. Assessment *as* learning, served as a subset of assessment *for* learning, is a process for students to develop metacognition, in which students actively participate in the learning process and reflect on their own learning to adjust or adapt their study with the support from instructors. Assessment *of* learning reflects the summative nature of assessment where instructors use assessment to provide a statement of students' proficiency that can be used by administrators to make decisions. Tools that permit the assessment *fo*r and *as* learning are conventionally labeled as formative assessment. Indeed, formative assessment is defined as an assessment activity that can provide evidence to be used as feedback for students to promote their learning and for instructors to modify their instruction (Black, Harrison, Lee et al., 2003; Black & Wiliam, 2009). Assessment *of* learning is conducted via summative assessment. Unlike formative assessment, summative assessment primarily serves the purpose of evaluating students' performance at the end of a unit, semester, or year (Emenike, Raker, & Holme, 2013).

Research has been conducted to understand instructors' view of the purpose of assessment and it is defined as "instructors' conception of assessment". Instructors' conceptions of assessment reflected instructors' affective attitude and beliefs about assessment that they espoused in their work, presumably in response to the policy as well as the classroom environments in which they work (Fulmer, Lee, & Tan, 2015). Browns (2006) developed a self-reported survey instrument, named as the Teacher Conceptions of Assessment inventory (TCoA), to characterize teachers' conceptions about assessment in primary schools. His study found that primary school teachers' conceptions can be interpreted through four different perspectives: 1) assessment as improvement of teaching and learning; 2) assessment as making schools and teachers accountable for their effectiveness; 3) assessment as making students accountable for their learning; 4) assessment is irrelevant to the life and work of teachers and students. The research group utilized their instrument across different countries (New Zealand, Cyprus, China, etc.) as well as school contexts (e.g., primary school, secondary school, teacher education, etc.) and found that both, culture and context, affected instructors' conceptions of assessment (G. T. Brown, Gebril, & Michaelides, 2019; G. T. Brown, Hui, Flora et al., 2011). Interestingly, Fletcher (2012) utilized the adapted TCoA to understand both instructors' and students' conceptions about assessment at four tertiary institutions. Within the 877 faculties and 1224 students who participated in the research, the majority of the faculties treated assessment as a way to facilitate their instruction and student learning. However, students treated assessment as irrelevant or overlooked it in the learning process. Reimann and Sadler (2017) interviewed 9 higher education instructors to understand their conceptions of assessment and associated practices by drawing a concept map. The results illustrated that variations were observed among participants in interpreting the purpose of assessment. Some instructors treated assessment as giving students feedback (assessment *for* learning, or formative assessment), others utilized assessment to assign grades and award degrees (assessment *of* learning, or summative assessment). Furthermore, they found that instructors' conceptions about assessment lead them to use a specific type of assessment, for example, instructors who believe assessment *for* learning incorporated reflective assessment in their classes, on the contrary, instructors who thought assessment *of* learning only utilized exams in their classes. To our knowledge, no research has been conducted to specifically explore general chemistry instructors' conception of assessment, since general chemistry serves as the gateway course for most STEM majors, and thus research is needed to understand this group of instructors.

Assessment practices

Assessment practices refer to instructors' use of assessment methods in a course which includes formal (e.g., exams, homework), and informal (e.g., Q&A during class time) ways (Fulmer, et al., 2015). Research has been conducted to understand what should be assessed and how to assess students in teaching chemistry.

In the disciplined-based education research community, the discussion has been made on what concepts should be assessed in chemistry courses. American Chemical Society Examinations Institute (ACS-EI) proposed the Anchoring Concepts Content Map (ACCM) (Holme & Murphy, 2012; Marek, Raker, Holme et al., 2018) to guide instructor to consider the "big ideas" that needed to be included in their assessment. Also, emphasis has been put on assessing students' conceptual understanding rather than rote memorization. Several ways have been proposed to assess students' conceptual understanding of specific concepts and identify misconceptions, such as, diagnostic test (Treagust, 1988) and concept inventories (Bretz, 2014; Mulford & Robinson, 2002). A more holistic approach has been proposed to guide instructors when design assessment has been advocated by Coopper (2019) named as Three-dimensional Learning. It suggests instructors to consider from three aspects when designing assessment which include: scientific and engineering practices, crosscutting concepts, and disciplinary core ideas. With the effort of discipline-based education researchers, resources have been developed based on Three-dimensional Learning Assessment Protocol (3D-LAP) to support instructors by using alternative ways to articulate their questions that capture students' conceptual understanding in a multiple choice format (Laverty, Underwood, Matz et al., 2016).

However, research demonstrated that the propagation of assessing students' conceptual understanding is not well adopted by STEM instructors. Momsen (2013) collected instructors' exams that have been utilized by two introductory science sequences, introductory biology, and introductory calculus-based physics, and analyzed their exam questions based on Bloom's taxonomy (1956) to identify the cognitive skills that have been required to answer each question. The results indicated that most of the STEM introductory courses focused on assessing students' ability to recall or solve algorithmic problems which only captured the lower-level cognitive skills. Similarly, Stowe (2017) analyzed 118 questions that have been assigned to students in organic chemistry courses across 4 elite universities by mapping on scientific and engineering practices. The results demonstrated that only 7% of the assessment items measured students' scientific and engineering practices, and the majority of the assessment items are focusing on recall basic content, solve algorithm problem or pattern recognition.

Less research has been conducted to characterize how to assess students in STEM courses. Erdmann (2020) investigated 42 instructors' assessment practice in teaching STEM disciplines at a Midwestern institution, the results demonstrated that 81% of the instructors used exams to assess students' learning, and only a small set of instructors used formative assessment during the learning process. More research is needed to understand what assessment tools have been utilized by instructors to get a better understanding of instructors' assessment practices.

Relationship between assessment practices and conceptions

Little research has been conducted to understand the connections between instructors' conceptions about assessment and their assessment practice. Postareff (2012) investigated 28 pharmacy teachers' conception of the purpose of assessment and their assessment practices. The results indicated that their conceptions are varied from "reproductive" which emphasized on recalling correct information, to "transformational" which emphasized on the development of students' thinking and understating. The results demonstrated that most of the instructors (n=22) held "reproductive" conceptions. They also found that instructors who endorsed "reproductive" conception are more likely to incorporate summative assessment, while formative assessment has been utilized by instructors who hold "transformational" conceptions. Offerdahl (2011) followed three biochemistry instructors for two years to understand their assessment thinking. She found that instructors' view of assessment changed as they were exposed to formative assessment activities (e.g., Clicker questions, reading questions), and instructors treated assessment as a way to understand students' learning or collect feedback from students after they use formative assessment activities rather than treated assessment as gathering information for the grades. However, the three instructors didn't incorporate the results collected from the formative assessment to reform their teaching which indicates a misalignment in instructors' conceptions and practices. The contradictory findings indicate more research is needed to untangle the complexity of instructors' conceptions about assessment and their associated practices.

Considering the conceptions of assessment are consciously or unconsciously affected instructors' assessment practices (Eley, 2006; Kember & Kwan, 2000), and there are less research having been conducted to investigate instructors' conceptions about assessment in higher education, especially in chemistry education, our research focused on exploring general chemistry instructors' rationales of utilizing assessment and their associated practices . General chemistry instructors were selected for this research since general chemistry severed as a gateway course for most of the STEM students. This research aims at understanding instructors' assessment practice and the rationale behind their assessment selection in teaching general chemistry. The research questions guiding this study are:

1. What are instructors' assessment practices in teaching general chemistry?
2. How do general chemistry instructors rationalize their assessment practices?
3. To what extent do general chemistry instructors' rationale for utilizing assessment related to their associated practices?

Methods

Participants

The participants were recruited from 14 different institutions across two states located on the East Coast of the United States. In total, 19 instructors consented to participate in the research. Their demographic information and the context of their teaching are summarized in Tables 4.1 and 4.2.

Data collection

The recruited instructors were asked to send their course artifacts, which include syllabus, exams, homework, as well as any assessment materials related to a general chemistry course taught during the academic year 2019-2020 (before COVID). After reviewing their course artifacts, instructors participated in a follow-up semi-structured interview, which focused on understanding instructors' assessment practices, their conceptions on the purpose of assessment, as well as explanations of their assessment practices (see appendix A2). Interviews lasted between 47 min and 93 min and were conducted through Zoom. All research was approved by the Institutional Review Board at the University of Virginia.

Table 4.1 Descriptive demographics for the participants

Demographic variables		**Number of participants**
Gender	Female	11
	Male	8
Academic Rank	Lecturer	7
	Professor	4
	Associate Professor	4
	Assistant Professor	3
	Visiting Professor	1
Teaching experience	Less than 5 years	5
	5.1 - 10 years	5
	10.1 – 15 years	4
	15 more years	5

Table 4.2 Descriptive context information for the participants

Context Variables		**Number of participants**
Class size	Less than 50	12
	more than 100	7
Classroom Layout	Fixed	13
	Flexible	5
	Both	1
Type of institution*	R1	6
	ML	4
	MM	3
	MS	1
	PUI	5

***** Based on Carnegie Classifications***: R1***: Doctoral Universities: Very High Research Activity; ***ML***: Master's Colleges & Universities: Larger Programs; ***MM***: Master's Colleges & Universities: Medium Programs; ***MS***: Master's Colleges & Universities: Small Programs; ***PUI***: Baccalaureate Colleges: Arts & Sciences Focus

Data analysis

Interviews were transcribed verbatim through Temi. Before the analysis, a number was assigned to each instructor and any information that could make the instructor identifiable were removed from the transcriptions. The researchers conducted a thematic analysis on the transcripts to identify instructors' assessment practices and their rationale for using assessment (Braun & Clarke, 2006). No preconceived themes or codes were used, rather the process was inductive (Saldaña, 2021) and facilitated by Nvivo (NVivo, 2000). Three researchers were involved in the coding process. Instructors' practices were coded by three coders, and the syllabus was used to confirm their articulation of their assessment practices. First, each of the coder memoed three interviews independently (Saldaña, 2021). Then, they met to discuss the memos and develop a consensus on how and what to memos. The three coders worked separately to generate a tentative codebook and met to discuss and developed a consensus codebook. After achieving agreement of the codebook, all three researchers coded one instructor separately and met to discuss disagreements and identify refinements needed to the codebook. This process was repeated with four more interviews. Two coders coded four participants (around 20% of the entire participants) separately with the refined codebook in order to establish interrater reliability, the selected instructors were different from the three instructors that used to design the coding book. The coders reached strong agreement as indicated by the Cohen's Kappa of 0.77 (Sim & Wright, 2005). The author of this dissertation coded the rest of the interviews. When uncertainties arose, she took detailed notes and discussed with the other two coders to resolve the issue. The rationale of using assessments were coded by two coders for all participants due to the complexity of capturing instructor reasoning in selecting assessment, and a consensus was made for all codes. The final codebook is provided in Appendix C15 and C16.

Results

Instructors' assessment practices

In this study, mid-terms and final exams (including group exams) were counted as summative assessment tools. Except for quizzes which count as both formative and summative due to instructors' ambiguous intentions of using them, other assessment tools are counted as formative assessment (Figure 4.1).

On average, instructors employed 6.2 ± 2.1 assessment tools in their general chemistry courses. Analysis of the interviews revealed that most instructors implement summative assessment tools. Indeed, almost two thirds of the assessment tools instructors use are summative, and this level of use was consistent across the instructors (M = 2.9 ± 0.5). All nineteen instructors assigned homework, mostly through online homework systems (e.g., Atkiv, ALEKS, Mind-Tap, etc.). Eighteen participants utilized mid-terms and cumulative final examinations. Quizzes were used by 13 participants, but their usage varied in frequency. Some (N=3) used quizzes every couple of weeks, others (N=5) weekly quizzes, and two of them used daily quizzes.

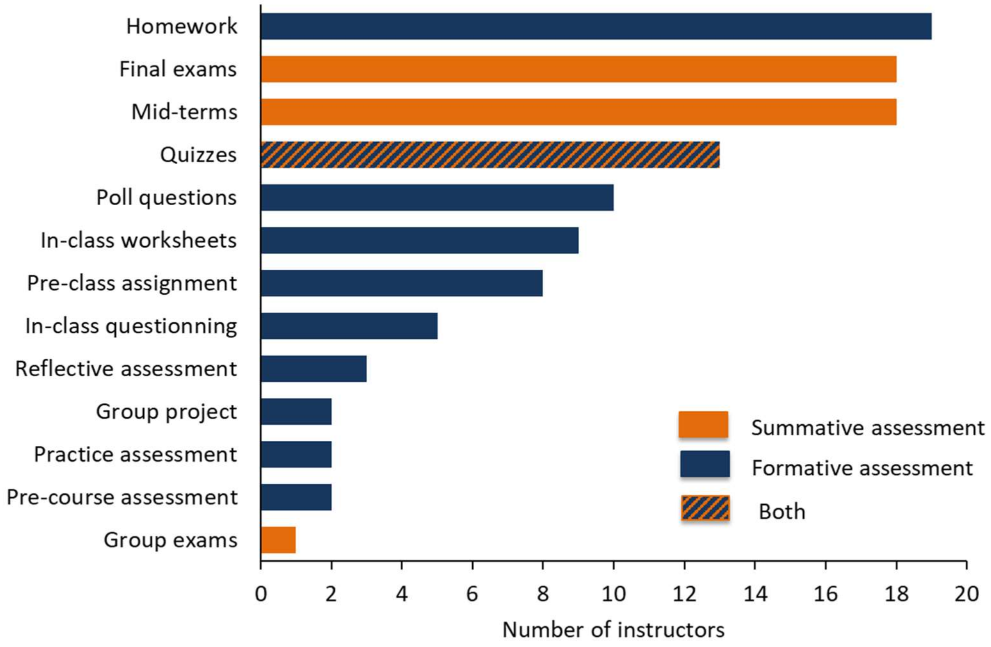


Figure 4.1 Summary of assessment practices of general chemistry instructors

Formative assessments were much less commonly used. On average, only a third of the assessment tools employed were of formative nature (M = 2.0 ± 1.4). Two instructors did not use any, and six only used one which represented 42% of the instructors. The most common formative tools were poll questions, in-class worksheets, and pre-class assignments. A quarter of the instructors (N=5) mentioned asking students questions during class, either formally to the whole class or while monitoring group work.

Two instructors used specification grading as an alternative way to assign grades for students (Nilson, 2015). One of these instructors (Instructor 11) still used final examination in the same way as the other instructors; however, instructor 5 used the final exam time as an opportunity for students to attempt to demonstrate mastery on any of the learning objectives that they did not pass during the semester. Both instructors quizzed their students on learning objectives addressed in one or two class sessions. Each quiz contains five to ten questions, and students' performance is based on a pass/fail grading system. If students do not pass the quiz, they have the opportunity to retake it either to re-schedule a time (for both instructors) or during the final time (for instructor 5). In both courses, the majority of the final grade is based on their performance for all the unit quizzes.

Instructors' views on the purpose of assessment

During the interview, instructors were asked to share their thinking about the purpose of assessment in general and to rationalize their specific assessment practices in their general chemistry course.

First, we asked the instructors to explain their beliefs about assessment and share their thoughts about the purpose and function of assessment in their course (Appendix C15). We leveraged Earl's assessment model (2006) to organize the purposes for assessment shared by the instructors: Assessment *for* learning, assessment *as* learning, and assessment *of* learning (Figure 4.2).

Almost all instructors (N=18; 95%) described a purpose for assessment that aligns with assessment *of* learning, i.e., assessments are used to provide evidence of student achievement. Most indicated that they use assessment to evaluate student learning (N=16; 84%). The following quotes provide examples of their description of this purpose:

> *"What do I believe assessment is ... assessment is a tool to see how well a student or a person, you know, depending on how general you're talking, how well they understand a concept, or how well they can perform a skill. So it's really a way to*

gauge where somebody is on a current, on some sort of goal that you have for them." Instructor 9

"[Assessment is] a statement of, uh, evaluation rather of the amount of material a student has learned relevant to the course." Instructor 15

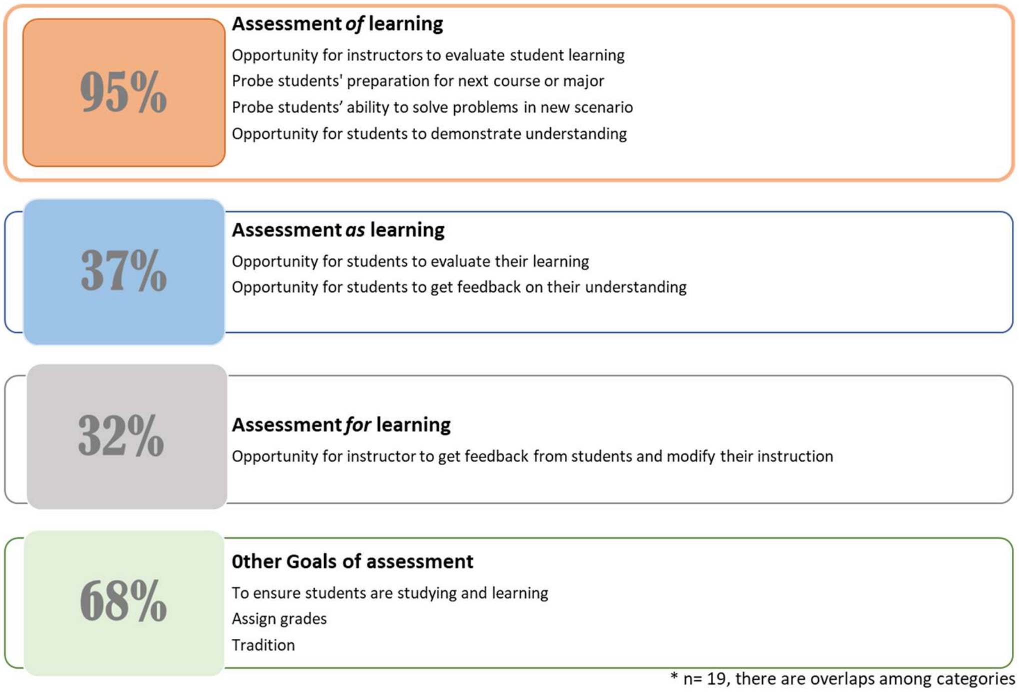

95%
Assessment *of* learning
Opportunity for instructors to evaluate student learning
Probe students' preparation for next course or major
Probe students' ability to solve problems in new scenario
Opportunity for students to demonstrate understanding
37%
Assessment *as* learning
Opportunity for students to evaluate their learning
Opportunity for students to get feedback on their understanding
32%
Assessment *for* learning
Opportunity for instructor to get feedback from students and modify their instruction
68%
Other Goals of assessment
To ensure students are studying and learning
Assign grades
Tradition
* n= 19, there are overlaps among categories

Figure 4.2 Summary of instructors' rational of utilizing assessment

Others more specifically described the purpose of assessment as a way to establish whether students are sufficiently prepared for their next course(s) (N=12; 63%):

"Most of them are taking another chemistry, even at least one more chemistry class. The assessment is a way, that's the , that's the way that I can be sure that they have gotten the foundational knowledge that they need to be successful in the next class. I mean, that's a prerequisite grades right? In order to take [the next] class, you have to have a C minus in the first class." Instructor 5

> *“Well, the function of assessment in general chemistry is … it really is an assessment of mastery. Uh, our goal is to make sure that students have mastered the learning objectives, or at least have a high enough level of competency within the learning objectives to move on to the, either the next section, the next unit in the course, but also to prepare them, um, that such that they have the tools and the competencies to move into the next level, which is our second year chemistry curriculum.”* Instructor 13

Two instructors (10%) indicated using assessment to probe students’ ability to solve problems in new scenarios: “*Whether or not they can take the concepts [we learned in] the class and then apply them to a new situation.”* Instructor 12

The next most common type of purpose was assessment *as* learning (N=7; 37%). Two different types of rationale fitted this category: opportunity for students to evaluate their learning (N=4; 21%) and opportunity for students to get feedback on their learning (N=5; 26%). As the following quotes indicate, the commonality of these two purposes is the focus on promoting student reflections:

> *“Opportunities to show the students where they're struggling, opportunities to help them start thinking about metacognition.”* Instructor 11
>
> *“The goal of the assessment is to provide them with feedback because students aren't always the best judges of how much they've learned and what their progresses are.”* Instructor 6

About a third of the instructors (N=6; 32%) felt that the role of assessment was to gather feedback on student learning in order to inform instruction, i.e., assessment *for* learning. The following quote exemplify this rationale:

> *"And so, if, if they're all doing terrible on something that's maybe indicative of, I'm not doing something well, um, or maybe they're not getting something. So, it's ideally a way to check their understanding so I can modify my methods so that they can get that understanding."* Instructor 14

Finally, 68% of the instructors mentioned other rationales. Some instructors (N=6; 32%) indicated that the purpose of assessment was to provide a grade to the students, something that they perceived as a requirement of the system:

> *"I think because … it has been communicated overtly and, um, and through non-verbal ways that I have to assess my students in one way or another and provide a letter grade."* Instructor 19

> *"Why I use assessment in general chemistry? So, part of it is expectations, right? Like I've got to have, I've got to have assessments in there because I've got to give them, I've got to give grades."* Instructor 12

Few pointed that giving assessment is a form of tradition, indicating again a sense of external pressure (N=5; 26%):

> *"Part of it is because that's the way that the courses have classically been, of course, that's definitely part of it."* Instructor 4

> *"And that's how I know how to do it. The way I set up my assessment. So, a series of tests, homework, grades, participation, lab, these are kind of a preset system. It's already kind of there. My hope is to kind of move away from it. But at the time I kind of was going with what other professors in my university had done, specifically with the other professor that was teaching, and I was kind of copying the structure."* Instructor 12

Finally, 21% of the instructors (N=4) stated that assessment is useful to ensure students are studying and engaging with the materials. The two quotes below illustrate that these faculty thought that assessment help motivate students to study:

> *"The assessment is something to scare the students. So, they study and learn the material."* Instructor 15

> *"If there were no assessment that contributed to the students' grade. There would be a lot of students who would struggle with motivation. For some students, it's the assessments themselves that motivate them to try and to learn and to participate. I mean, hopefully that's not the case for all students, hopefully, many students are motivated because they're intrinsically interested in what they're doing but that's not going to be the case for everybody."* Instructor 6

To conclude, most of instructors in this study view assessment as a mean to evaluate the level of learning that students have achieved at a particular time point. Indeed, 84% the participants exclusively mentioned that purpose. Fewer instructors mentioned the potential for assessments to support student learning and instructional effectiveness.

Rationale for utilizing specific types of assessment

After describing the different assessment tools instructors were using in their general chemistry course, we asked them to share their reasoning for using each of these tools. We classified the rationales they provided (see Appendix C17) into the Earl's assessment model. It is important to note that not all instructors commented on each tool they were using. For example, only two of the nine instructors who assigned a pre-class assignment provided a rationale for utilizing it, and about a third of the instructors did not explain why they assigned homework. Consequently, we do not have a reasoning for each instructor for each tool they employed. We report here on the tools that were commented on the most.

Except for homework, which were mostly seen as tools to help students monitor their understanding, most of the reasoning that faculty provided did not fall into Earl's assessment model (Table 4.3). The most common reasons advanced by the instructors for using final exams and mid-terms was that it was tradition (N=9, 47%):

> *"Part of it is tradition. We've always traditionally, at this university and all universities I have been, have had some form of, uh, assessment with, with respect to midterms and final exams."* Instructor 13

The use of final exams and mid-terms is so embedded in the fabric of traditional course design that it did not occur to instructors to consider something different:

> *"I guess exams [mid-terms] … I've always, we've always used exams since when I started teaching and I guess I, that's one of those things I don't think I've ever really questioned like no it's chemistry, we're gonna give exams."* Instructor 11

> *"I don't think I thought very much about exams and a final. It's just that's what everybody does so I am doing them too. I don't think that was really something I put a lot of thought into."* Instructor 6

Table 4.3 Types of rationale general chemistry instructors provided for utilizing specific types of assessment, and the number of instructors who provided a rationale for each type of assessment is provided in parenthesis

Type of assessment	**Purpose of assessment**			
	Assessment *of* learning	**Assessment *for* learning**	**Assessment *as* learning**	**Other**
Final exams (N=12)	4	0	0	10
Mid-terms (N=16)	5	1	6	13
Homework (N=13)	3	1	10	4
Quizzes (N=8)	4	0	1	6
In-class assignments* (N= 7)	0	2	2	6

** include poll questions, in-class worksheets, and in-class questioning*

With respect to quizzes, several instructors (N=4) indicated using them in order to ensure that students are studying and processing the materials:

> *"Then the quizzes and tests are like to encourage them to study material before too much time passes by, to forget it. So, a lot of it is like to make sure that they're staying on top of the material, and then move on from there."* Instructor 2

Finally, instructors used in-class assignments to provide opportunities for students to talk to each other (N = 4, 57%), and to get them engaged with the material (N = 2, 29%).

Reasons for using a particular type of assessment that could be classified under Earl's assessment model mostly felt into the Assessment *of* learning and Assessment *as* learning categories (Table 4.3). For example, final exams were used as Assessment *of* learning, while homework were mostly used to help students monitor their learning (Assessment *as* learning). Mid-terms were perceived to be useful for both of these reasons. The rationale for using In-class assignments had only two purposed aligned with the Earl's model, and others treat it as an opportunity for students to discuss with each other, or boost their grades.

Few instructors provided rationale that felt into the Assessment *for* learning indicating that instructors are not leveraging the feedback on student to inform their teaching practices.

Relationship between instructors' views on the purpose of assessment and their instructional practices

In this section, we explore one relationship identified in the TCSR model, the connection between Teacher Thinking (i.e., the purpose general chemistry instructors described for using assessment) and Instructional Practices (i.e., the types of assessments these instructors use). Focusing solely on the three purposes of assessment (assessment *for*/*as*/*of* learning), we explored the extent to which instructors identified rationales preferentially for one or more of these purposes. This analysis led to the identification of three groups: 1) instructors who indicated that assessment served all three purposes (N=3; 16%); 2) instructors who indicated that assessment served assessment *of* learning and either assessment *as* or *for* learning (N=7; 37%); and 3) instructors who only mentioned rationales that fitted the assessment *of* learning (N=8; 42%). There is one instructor whose rationale for using assessment can't be categorized into Earl's model, and they only use assessment to ensure students are studying and learning. In the following discussion, emphasis was made on the three groups that could be categorized into the Earl's model.

We tabulated the type of assessments used by instructors in the three different group and observed some alignment between instructors' description of their purpose for assessing and their practices. For example, we expected that instructors identifying all three types of purposes would employ a mixture of formative and summative assessment tools and thus would have the highest number of different tools. Table 4.4 demonstrated an increase in the average number of assessment types utilized as the number of purposes for assessing increase. Instructors who felt into the assessment *of/as/for* learning group employed the most, while those who only identified purposes aligned with Assessment *of* learning employed the least. However, we expected that this latter group would rely solely on summative tools. However, more than half of the instructors in this group (Assessment *of* learning) also used formative assessment tools such as in-class and pre-class assignments (Table 4.5).

While the number of summative assessment tools was very consistent across and within all three groups, the number of formative assessment tools was quite variable, especially within groups. For example, within the group of Assessment *of/as/for* learning, two of the three faculty employed three different types of formative assessment, one instructor only relied on in-class worksheets.

Another aspect of instructors' assessment practices that was captured in the interviews is their usage of the feedback provided by the assessments (Figure 4.3). Specifically, we asked instructors to describe the type of feedback that each assessment tool provided to them and how they leveraged this feedback. Figure 4.3 shows that feedback was used by most of the instructors to modify their instructional practices. However, differences between the three groups of instructors were observed.

Table 4.4 Average number of assessment tools employed by each group of instructors

Type of assessment	**Type of instructor's views on the purpose of assessment**		
	Assessment *of* learning (n = 8)	**Assessment *of* and *as/for* learning (n = 7)**	**Assessment *of, as*, and *for* learning (n = 3)**
Average total number of assessment tools	5.3 ± 2.2	6.1 ± 1.8	6.7 ± 0.6
Average number of summative assessment tools*	2.8 ± 0.7	3.1 ± 0.4	3.0 ± 0.0
Average number of formative assessment tools*	1.5 ± 1.2	2.1 ± 1.3	2.3 ± 1.2

** Quizzes were not included in these categories since they were used for different purposes by instructors in our sample*

Table 4.5 Categorization of Instructors conceptions about assessment and their associated assessment practices

Type of assessment	**Type of instructor's views on the purpose of assessment**		
	Assessment *of* learning (n = 8)	**Assessment *of* and *as/for* learning (n = 7)**	**Assessment *of, as*, and *for* learning (n = 3)**
Final exams	87%	100%	100%
Mid-terms	87%	100%	100%
Homework	100%	100%	100%
Quizzes	75%	71%	100%
In-class assignments*	62%	100%	100%
Pre-class assignments	50%	43%	67%
Reflective assessment	0%	14%	33%

** include polling questions, in-class worksheets, and in-class questioning*

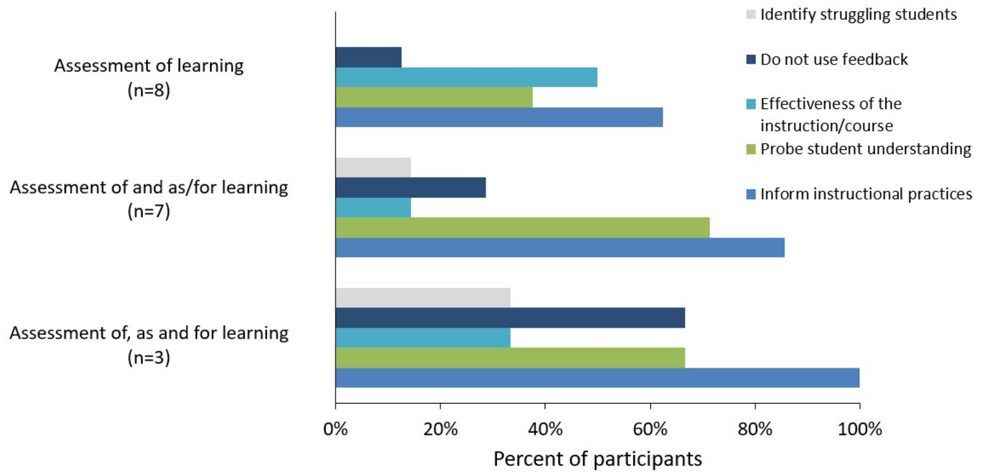


Figure 4.3 Instructors' usage of feedback received from assessment

Over 80% of the faculty in the groups who identified more than two purposes of assessment highlighted at least two different purposes for assessing leverage assessment feedback to modify their instruction while only 63% of the Assessment *of* learning group mentioned it. Two instructors in this latter group (n = 2; 33%) indicated using the feedback to inform their practices in future iterations of the course, while all instructors in the other two groups use the feedback to inform practices in their current courses. The following two quotes illustrate this contrast:

> *" For exam material, if I see that things are missed a lot, most likely, I don't go back and cover it unless it's something that's gonna build. However, I do actually make a note for the next time I teach the course, and remember the topics that they struggled with."* Instructor 7 – Assessment *of* learning group

> *"Since I'm grading my own exams as well um I will look to see, so, multiple choice questions, if one particular question was completely missed by everyone, and then I will go over that."* Instructor 8 – Assessment of and *as/for* learning group

All three groups indicated gathering feedback from a variety of assessment tools in order to inform their practices. However, the Assessment *of* learning group mentioned exams and quizzes more than the other groups; the other two groups relied more on in-class assignments.

All three groups indicated using assessment to probe students understanding, although the Assessment *of* learning group mentioned this aspect the least (n = 3; 38%). Half of the instructors in this group tended to leverage feedback from assessment to evaluate their instructional effectiveness or the effectiveness of the course, compared to less than a third of the instructors in the other two groups.

> *"Now because of the way that we write exams, and in particular, the way we grade exams … because our exams are written by learning objective and there's … an automated way of grading, we can look at how every single student did on every single question. So,*

> *we can, if we know that question 10 focuses on these three learning objectives, we can then look at the data and say, okay, 85% of the students got question 10 right. We're doing the right thing there."* Instructor 13 – Assessment *of* learning group

Finally, the analysis of the data indicated that homework is seldom looked at by instructors to gather feedback. Only five instructors out of the nineteen mentioned homework when responding to the question about their use of the feedback provided by assessment. Three of these five instructors indicated specifically never using homework as a feedback mechanism. The following quote illustrates this point:

> *"I honestly don't have time to look at the homework system. There's just so much data there and it's presented in, I have never found a useful way to get stuff out and more, more often than not the students are gaming the system. So I don't feel like it gives me good feedback."* Instructor 18 – Assessment *of/as/for* learning

Overall, instructors who provided rationales across all three purposes presented in Earl's assessment model, incorporated more types of formative assessment compared to instructors who identified rationales along one or two purposes only. However, instructors whose rationales only aligned with Assessment *of* learning still employed formative assessment tools. Moreover, this group, unexpectedly, leveraged the feedback provided by assessment to inform their instructional practices. These results indicate that some instructors within the Assessment *of* learning group did not fully articulate rationales for assessing that aligned with their practices. Rationales and practices where better aligned for instructors in the other two groups.

Discussion

This research aims at understanding instructors' rationale for using assessment in general chemistry and their associated practices.

The results demonstrated that summative assessment (e.g., cumulative finals, mid-terms) is predominantly utilized by all participants to assess students in general chemistry which is aligned with the research that has been conducted a decade ago (Momsen, et al., 2013). Formative assessment tools are much less used by instructors and variability has been observed for how many formative assessment tools have been used both within and across the three identified groups (Assessment *of* learning, Assessment *of* and as/for learning, Assessment *of*, *as*, and *for* learning). Only three instructors indicated using assessment to promote students' reflection, such as writing "muddiest point" at the end of the class to identify their confusion about the instruction which resonates with previous research (Raker, Emenike, & Holme, 2013) that instructors are lack understanding of formative assessment. The results indicated a need to educate faculty about formative assessment and the different types of formative assessment tools that could be utilized to promote students' learning.

Variations have been observed from instructors in their rationale for utilizing assessment in general chemistry. By leveraging Earl's purposes of assessment model (Earl & Katz, 2006), assessment *of* learning was mentioned by all but one instructors. Most of the instructors are interested in evaluating students' learning (n= 16, 84%), and ensuring students are prepared for their next level of study (n=12; 63%). Much fewer instructors recognize the purpose of assessment *as* learning (build students' metacognition abilities), or *for* learning (utilize assessment to inform instruction), and an instructor was able to identify more than one aspect of assessment (*for, of*, *as*, learning).

Other purposes of assessment have been identified by instructors, such as providing grades, following tradition, or ensuring students are studying and learning that cannot be categorized in the Earl's model. The majority of instructors' rationales for utilizing exams were classified under tradition which indicated most of them followed the traditional way of what has always been done in assessing students in general chemistry, and they feel pressure to assign a grade to students at the end of the semester. There were only two instructors who utilized a different assessment framework, specification grading

(Nilson, 2015), and the results indicated a need to elevate the propagation of assessment to the same level as instructional practices when conducting professional development for faculties to elicit their awareness of the existence of different types of assessment, other than just following the tradition.

Instructors' rationales of using each specific type of assessment (e.g., homework, mid-term, finals) have been investigated in this research, and few rationales could be categorized into the Earl's model, and only a small part of the rationales classified in assessment *for* learning. However, when we prompted their usage of feedback for these specific assessments, most of the instructors indicated that they used the feedback to modify their instructional practice which is not aligned with their articulation of reasoning of utilizing specific type of assessment.

Regarding their rationale for using assessment and associated practices, the two groups that identified more than two purposes within the Earl's model demonstrate a better alignment that resonates with the research (Postareff, et al., 2012) that instructors' conceptions will affect their associated practices. However, around half of the group who only recognized the assessment *of* learning still used formative assessments and they indicated leveraging the feedback to inform their instruction, even though they used less formative assessment than the other two groups. This indicated the complex relationship between the instructors' conceptions and practice (Offerdahl & Tomanek, 2011), which emphasized the importance of considering instructors' practices when exploring instructors' conceptions. Exploring both practices and conceptions of assessment concurrently provides a more accurate, in-depth characterization of their thinking about assessment.

Lastly, homework was assigned by all participants through an online Homework system. Around half of the instructors' rationale for assigning homework served as a formative purpose , such as, an opportunity for students to get feedback on their learning or an opportunity for students to relearn things that they did not understand. The rest of the instructors who assigned homework used it to evaluate students' learning and it is not clear if any feedback has been given to students during or after students finished their homework, and some instructors only use homework to boost students' grades. On the other side, instructors ignored that feedback from the homework could promote their instructional effectiveness since none of the participants pointed out the assessment *for* learning in their rationale for assigning homework. The majority of the instructors relied on feedback generated from the online homework system to guide students' learning, while the effectiveness of feedback that has been given is unclear and further research is needed to evaluate its validity.

Conclusions

This study aims at understanding instructors' assessment practices in general chemistry and the rationales associated with their practices. The results indicated that summative assessments are predominately utilized by participating instructors. Investigation of instructors' reasoning in selecting a specific type of assessment indicates a lack of using formative assessment to modify their instructional practices. By leveraging Earl's purpose of assessment model, three groups of instructors were identified based on their thinking about the purpose of assessment. Analysis of their practices demonstrated that the group who recognized all three aspects of assessment (*for, as, of* learning), or at least two purposes of assessment (*of*, and either *for* or *as)* incorporated more formative assessment in their practice which indicted an alignment between their conceptions and practices. However, the group who only recognized assessment *of* learning still incorporated some formative assessment in their practices. By exploring instructors' usage of feedback from assessment, the results indicated the group who only recognize the purpose of assessment is *of* learning used feedback from their assessments to inform their instructional practices which indicated their rationale of using assessment are not fully aligned with their practices which highlights the complexity in exploring rationale of using assessment and associated practices.

Limitations

There were nineteen general chemistry instructors participated in this exploratory study in various types of institutions on the east coast of the United States. However, the sample size may limit the results to be inferred to general conclusions for all general chemistry instructors across the country. Some of the instructors didn't articulate their reasoning in utilizing a specific type of assessment may be due to the lack of some follow-up questions in probing their rationale. Future research could be done to focus on understanding instructors' rationale for selecting a specific type of assessment with expanded sample size.

Implications

This exploratory study attempted to understand instructors' conceptions about assessment and their assessment practices in teaching general chemistry at the postsecondary level. The results provide evidence for researchers that the current stage of innovation of assessment is still lacking with the effort has been investigated for decades. However, new ways have been incorporated by some of the instructors to evaluate their assessment (e.g., specification grading), and poll questions have been used by more than half of the participants, indicating their awareness of pedagogical/assessment innovation. The findings could be used as a resource in professional development activities to assist instructors to reflect on their assessment practices and discuss better ways to assess students in general chemistry.

References

Black, P.,Harrison, C.,Lee, C., et al. (2003). Formative and summative assessment: Can they serve learning together. *AERA Chicago, 23*.

Black, P.,Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education), 21*(1), 5-31.

Bloom, B. S. (1956). Taxonomy of educational objectives: The classification of educational goals. *Cognitive domain*.

Braun, V.,Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

Bretz, S. L. (2014). Designing assessment tools to measure students' conceptual knowledge of chemistry. In *Tools of chemistry education research* (pp. 155-168): ACS Publications.

Brown, G. T. (2006). Teachers' conceptions of assessment: Validation of an abridged version. *Psychological reports, 99*(1), 166-170.

Brown, G. T.,Gebril, A.,Michaelides, M. P. (2019). *Teachers' conceptions of assessment: A global phenomenon or a global localism.* Paper presented at the Frontiers in Education.

Brown, G. T.,Hui, S. K.,Flora, W., et al. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research, 50*(5-6), 307-320.

Earl, L. M.,Katz, M. S. (2006). *Rethinking classroom assessment with purpose in mind: Assessment for learning, assessment as learning, assessment of learning*: Manitoba Education, Citizenship & Youth.

Eley, M. G. (2006). Teachers' conceptions of teaching, and the making of specific decisions in planning to teach. *Higher Education, 51*(2), 191-214.

Emenike, M.,Raker, J. R.,Holme, T. (2013). Validating chemistry faculty members' self-reported familiarity with assessment terminology. *Journal of Chemical Education, 90*(9), 1130-1136.

Erdmann, R.,Miller, K.,Stains, M. (2020). Exploring STEM postsecondary instructors' accounts of instructional planning and revisions. *International Journal of STEM Education, 7*(1), 1-17.

Fletcher, R. B.,Meyer, L. H.,Anderson, H., et al. (2012). Faculty and students conceptions of assessment in higher education. *Higher Education, 64*(1), 119-133.

Fulmer, G. W.,Lee, I. C.,Tan, K. H. (2015). Multi-level model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy & Practice, 22*(4), 475-494.

Gijbels, D.,Segers, M.,Struyf, E. (2008). Constructivist learning environments and the (im) possibility to change students' perceptions of assessment demands and approaches to learning. *Instructional Science, 36*(5), 431-443.

Hanauer, D. I.,Bauerle, C. (2012). Facilitating Innovation in Science Education through Assessment Reform. *Liberal education, 98*(3), 34-41.

Holme, T.,Murphy, K. (2012). The ACS Exams Institute undergraduate chemistry anchoring concepts content map I: General Chemistry. *Journal of Chemical Education, 89*(6), 721-723.

Kember, D.,Kwan, K.-P. (2000). Lecturers' approaches to teaching and their relationship to conceptions of good teaching. *Instructional Science, 28*(5), 469-490.

Laverty, J. T.,Underwood, S. M.,Matz, R. L., et al. (2016). Characterizing college science assessments: The three-dimensional learning assessment protocol. *PloS one, 11*(9), e0162333.

Marek, K. A.,Raker, J. R.,Holme, T. A., et al. (2018). The ACS Exams Institute undergraduate chemistry anchoring concepts content map III: inorganic chemistry. *Journal of Chemical Education, 95*(2), 233-237.

Momsen, J.,Offerdahl, E.,Kryjevskaia, M., et al. (2013). Using assessments to investigate and compare the nature of learning in undergraduate science courses. *CBE—Life Sciences Education, 12*(2), 239-249.

Mulford, D. R.,Robinson, W. R. (2002). An inventory for alternate conceptions among first-semester general chemistry students. *Journal of Chemical Education, 79*(6), 739.

Nilson, L. B. (2015). *Specifications grading: Restoring rigor, motivating students, and saving faculty time*: Stylus Publishing, LLC.

NVivo, Q. (2000). Qualitative data analysis program. *QSR International Pty Ltd, Melbourne, Australia*.

Offerdahl, E. G.,Tomanek, D. (2011). Changes in instructors' assessment thinking related to experimentation with new strategies. *Assessment & Evaluation in Higher Education, 36*(7), 781-795.

Postareff, L.,Virtanen, V.,Katajavuori, N., et al. (2012). Academics' conceptions of assessment and their assessment practices. *Studies in Educational Evaluation, 38*(3-4), 84-92.

Prosser, M.,Trigwell, K. (1999). *Understanding learning and teaching: The experience in higher education*: McGraw-Hill Education (UK).

Raker, J. R.,Emenike, M. E.,Holme, T. A. (2013). Using structural equation modeling to understand chemistry faculty familiarity of assessment terminology: Results from a national survey. *Journal of Chemical Education, 90*(8), 981-987.

Reimann, N.,Sadler, I. (2017). Personal understanding of assessment and the link to assessment practice: the perspectives of higher education staff. *Assessment & Evaluation in Higher Education, 42*(5), 724-736.

Saldaña, J. (2021). *The coding manual for qualitative researchers*: sage.

Sim, J.,Wright, C. C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical therapy, 85*(3), 257-268.

Stowe, R. L.,Cooper, M. M. (2017). Practicing what we preach: assessing "critical thinking" in organic chemistry. *Journal of Chemical Education, 94*(12), 1852-1859.

Stowe, R. L.,Cooper, M. M. (2019). Assessment in chemistry education. *Israel Journal of Chemistry, 59*(6-7), 598-607.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International journal of science education, 10*(2), 159-169.

Trigwell, K.,Prosser, M. (1991). Improving the quality of student learning: the influence of learning context and student approaches to learning on learning outcomes. *Higher Education, 22*(3), 251-266.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*(1), 3-14.

Zimmerman, B. J.,Bandura, A.,Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal, 29*(3), 663-676.

Chapter 5 - Conclusions, Implications, and Future Directions

This dissertation aims to characterize three different factors represented in the Teacher-Centered Systemic Reform (TCSR) model to inform future instructional reform projects.

In Chapter 2, we explored the characteristics of departmental climate around teaching and the relationship between climate and use of evidence-based instructional practices. Indeed, departmental norms, policies and practices are often identified in the literature as barriers to instructional innovation. The analysis of the data gathered during the development and implementation of the Departmental Climate around Teaching (DCaT) survey (a new instrument that was developed as part of this work) points to difficulties in measuring departmental collective climate around teaching since most elements that are used to define a climate (such as, policies, expectations, or practices) are lacking when it comes to teaching. The analysis also did not reveal a relationship between psychological collective climate and instructors' uptake of learner-centered instructional practices. This work highlights the need for departments to establish expectations, policies, and practices around teaching similar to what is done in research. Without these elements, instructional innovation at the department level is unlikely to be enacted. The product of this work, the DCaT, provides a means for departments and entities engaged in departmental reform efforts to identify a baseline and monitor progress in establishing a psychological collective climate around teaching. Moreover, the instrument provides some actionable constructs that they can be leveraged to inform the growth of a supportive environment for teaching and learning. The future direction of this project will focus on measuring departmental collective climate through an institution that has been successfully incorporated teaching innovation, such as Michigan State University innovated its general Chemistry curriculum by incorporating learner-centered instructional practices named Chemistry, Life, the Universe, and Everything (CLUE) (Cooper & Klymkowsky, 2013). Doing a case study to understand their departmental climate around teaching will shed light on how the efforts have been invested to support their instructors for teaching innovation and potentially identify the constructs that could be used to measure departmental collective climate.

In Chapter 3, we aimed to contribute to the current conversations about the role that teaching evaluation could play as a driver for instructional change. We were particularly interested on the recommendations to use several sources of evidence in these processes. Recognizing the challenges in parsing out and selecting a set of instruments to capture various aspects of teaching practice, we were interested in characterizing the complementarity of two instruments, the Classroom Observation Protocol for Undergraduate STEM (COPUS) and the Learner-centered Teaching Rubrics (LCTR). COPUS has been adopted extensively for capturing instructors' in-class activities, and LCTR as a holistic rubric has been used to capture instructional practices for both in-class and out-of-class. The results demonstrate a partial misalignment between the two instruments. While they both aligned when characterizing instructors' course instructional styles (e.g., mostly lecturer versus mostly student-centered), their characterizations of the nuances of instructional practices were not congruent. These findings display the challenges in measuring instructional effectiveness and the need to carefully select evaluation instruments that align with and can inform desired outcomes. This work also leads to a call for more research on 1) the complementarity of instruments that have been developed to measure instructional practices and 2) best practices for evaluating instructional quality based on the triangulation of evidence collected across a set of instruments. The identification of a set of measures that provide valid and reliable assessment of instructional quality is essential to enhance the credibility of teaching evaluation and to ensure their role as drivers of instructional innovations.

One future direction of this project could be integrating different instruments into one framework to holistically capture instructional practices. For example, combining LCTR and COPUS as one instrument then valid the instrument through different perspectives to generate a new measurement tool. As formal research in measuring instructors' uptake of Scientific Teaching (ST), the research group developed a survey instrument named Measurement Instrument for Scientific Teaching- Observable (MISTO). They triangulated the data that has been collected through instructors’ self-reported usage of ST, students' evaluation of instructors' use of ST, and external observers’ observations of instructors' teaching practices (Durham, et al., 2018). Triangulating data through different perspectives will provide some insights in measuring teaching effectiveness and find the common constructs that could be used to collect credible evidence.

Finally, in Chapter 4, we explored general chemistry instructors’ assessment practices and their rationales for utilizing assessment. This characterization is crucial as the chemistry education research community is advocating for the implementation of new assessment approaches. Understanding instructors’ approaches and thought processes with respect to assessment could inform these reform efforts. The results of interviews conducted with nineteen general chemistry instructors indicate that summative assessment tools (e.g., mid-term, cumulative finals) are prevalent in assessing students, and their rationale for using assessment is more toward assessment *of* learning with the Earl’s assessment model (Earl & Katz, 2006). Three groups of instructors have been categorized into who only recognized the purpose of assessment is *of* learning, who identified assessment *of* learning, and *for* or *as* learning, and who recognized all three purposes of assessment. The group who identified all three purposes of assessment used more types of assessment than other groups, however, instructors who only recognized assessment *of* learning still utilized formative assessments which indicated the complexity of understanding their conceptions of assessment and associated practices. Instructors’ rationale of utilizing specific types of assessment has been investigated and not all instructors were able to articulate the reasoning behind their usage. The instructors who provided rationales for using a specific type of assessment ignored that feedback from the assessment could be served as a way to modify their instructional practices. However, further investigation of their usage of feedback illustrated that instructors leveraged the feedback provided by assessment to inform their teaching. The results indicate that some misalignment has been observed in the group of instructors who only recognized assessment *of* learning in their conceptions about assessment and associated practices.

The findings presented here focus on a subset of the data that was collected. Other aspects of the interview that remain to be analyzed include the exploration of 1) factors influencing instructors' assessment practices, 2) the relationship between instructors' conceptions of teaching and learning and their assessment practices, and 3) their thought processes when selecting particular questions to be included in different assessment tools (e.g, homework, in-class activity, exams). The work conducted to date emphasizes the need for reform efforts (i.e., center for teaching and learning, pedagogical program run by professional organizations) to provide training on assessment to future and current STEM instructors. For years, the focus has been on enhancing their pedagogical toolbox, but the results have been mixed. Providing tools for instructors to implement assessment practices that will provide them with more insights into the level and quality of student understanding, in turn, will encourage them to explore and potentially implement new instructional strategies.

A theme that runs across all three chapters was the complexity of characterizing the various factors represented in the Teacher-Centered Systemic Reform (TCSR) model and some of the relationships among these factors. This body of work calls for more qualitative and quantitative investigations of the factors that are hypothesized to inform STEM instructors' conceptions of teaching and learning and their practices. This work is essential to enhance the effectiveness of instructional reform efforts.

References

Cooper, M.,Klymkowsky, M. (2013). Chemistry, life, the universe, and everything: A new approach to general chemistry, and a model for curriculum reform. *Journal of Chemical Education, 90*(9), 1116-1122.

Durham, M. F.,Knight, J. K.,Bremers, E. K., et al. (2018). Student, instructor, and observer agreement regarding frequencies of scientific teaching practices using the Measurement Instrument for Scientific Teaching-Observable (MISTO). *International Journal of STEM Education, 5*(1), 1-15.

Earl, L. M.,Katz, M. S. (2006). *Rethinking classroom assessment with purpose in mind: Assessment for learning, assessment as learning, assessment of learning*: Manitoba Education, Citizenship & Youth.

Appendix

Appendix A. Interview Protocols

A1. Cognitive interview protocol

Thank you for accepting the invitation to be interviewed for this research project. Have you had an opportunity to read the consent form? Can/Have you signed the consent form?

We would like to audio record this interview. Is that okay with you? Ok, I have begun recording and I'll ask again for the record, is it okay that I record this conversation?

As a reminder, we will be taking measures to assure your anonymity. Any publications using results from this interview will be published completely anonymously, including removing your name, other names you mention, your institution, or anything that makes your quotes identifiable.

I'm going to ask you some questions related to the climate survey that you filled out. Our research project aims to understand your view about your departmental climate with respect to teaching. As a result, please answer these questions honestly and without consideration of "answers that you think I'm looking for." There is not right or wrong answer. Your opinion is what we are interested in.

The goal of this interview series is to help us establish the validity of the survey: we want to ensure that the questions are clear to our interviewees, that they have the same meaning across different interviewees, and that this meaning matches what we were trying to investigate.

I'll be taking notes as you talk, don't feel like you need to slow down for me; I have the recorder to help me ensure I capture everything you say. Also, I try to avoid talking about the same thing you talked about earlier, but addressing seemingly similar points may be a part of the research process.

Do you have any questions before we begin?

Now let's turn to the survey you filled out a couple of days ago. Here is a copy of the survey answer; it can help you recall the questions and your responses. Again the goal of this interview is to help us know if the survey works as we intend. Could we start?

1. How comfortable did you feel answering these questions?
2. Were there items that were difficult to understand? Which one(s) and why?
3. Were there items that you found difficult to answer? Which one(s) and why?
4. Were the response options provided clear and appropriate for the question?
5. This questionnaire was about departmental climate around teaching.
    a. Was there anything not included in the survey that are important to you and that you think should be included?
    b. Were there items that you feel do not fit with that framework?
    c. (if not already addressed before) Were there items that did not apply to your environment?

Now I'm going to transition to the second section of today's interview. Based on your answers, we have the following questions:

1. You answered "neither agree or disagree" in these items - we highlight these items in "YELLOW" in the survey.
   What message did you mean to send by checking this option?

2. We noticed some inconsistencies among items that we thought measured the same idea. We want to know what you were thinking when you answered those questions. We highlight this item in "BLUE" in the survey.

3. Last question, do you feel that the demographic questions were appropriate?

Thanks for taking the time to participate in this interview!

A2. Interview Protocol of assessing instructors' conceptions about assessment and associated practices

Thank you for accepting the invitation to be interviewed for this research project. Have you had an opportunity to read the consent form? Can/Have you signed the consent form?/I already see you sign the concept form, so we are good to go!

This research aims at understanding instructors' beliefs and practice about assessment, and reasoning to select student' assessment practice.

- We would like to audio record this interview. Is that okay with you? Ok, I have begun recording and I'll ask again for the record, is it okay that I record this conversation?
- As a reminder, any publications using results from this interview will be published completely anonymously, including removing your name, other names you mention, your institution, or anything that makes your quotes identifiable.
- Do you have any question before we start?

Interview Protocol

Course information and teaching experience:

1. How long have you been teaching **Gen chem**?
    1) What is the structure of the course? (lecture, lecture + lab)
    2) How many students are enrolled in this course?
    3) What is the classroom layout? (amphitheater, round table, moveable desks)
    4) Do you have access to technologies and tools that could be used in the class (e.g. Clickers, whiteboard)
    5) Did you co-teach this course?
        a. If YES: Do the instructors design the assessment together? Which one (homework exams, clicker questions)? How?
2. Can you walk me through a typical day in your class in general chemistry?

Now, we want to understand your thinking about teaching and learning in general chemistry. In order to do that, we have a set of metaphors to help us understand your beliefs…

We want to you to consider from two perspective, the first is from Philosophy view which means what you want your general chemistry class to be, and another is from the practice view which means, what is your general chemistry class looks like in real situation.

Beliefs Questions:

3. Here are a set of metaphors about your thoughts on teaching and learning: which one of these aligns best with your thinking?
    1) Teaching is guiding: I see myself leading my students on a treasure hunt. I have a map that shows us the way. Sometimes the path is hard and some-times it is easy, but it is always worth it when we get to the end.
    2) Teaching is nurturing: It is a sunny day. I see myself holding a watering can and carefully attending to my seedlings. I make sure that the soil, water, and climate are rich and right for each seedling so that each will develop and blossom.
    3) Teaching is molding: I am seated at a potter's wheel with a lump of clay. I carefully mold the clay into a well shaped and beautiful vase. Sometimes it takes pushing and prodding to get the vase to develop.

    4) Teaching is transmitting: I have a large sum of money, which I deposit into a series of accounts. The goal is to deposit as much money as I can into each account so that each account has a high balance.
    5) Teaching is providing tools: I wear a large tool belt. As each worker constructs his or her house, I provide the builder with the tools he or she will need to be successful in completing the project.
    6) Teaching is engaging in community: I am part of a community that is building a house. We collectively decided that we need a house and then we design and build it together.
4. what do you believe assessment is?
5. What is the function of assessment in general chemistry?
6. Why did you use assessment in general chemistry?

Macro/meso-level:

7. Do you experience external influences that affect the way you assess your students in general chemistry course? How?
    a. Follow-up with whatever is not addressed in the answer:
        i. National level like ACS exams
        ii. Institutional level –
            - the institution is forcing them to assess in a particular way(institution policy, force instructor to do a certain thing. Institution force a certain learning goals)
            - Other departments at your institution influence assessment because many students from their departments are enrolled in your course
        iii. Students ~~:~~Did you feel pressure from the students to assess in a particular way or do specific types of assessment?

Assessment Practice Questions:

8. How do you assess your students in general chemistry?
    1) What assessments did you use?
    Or “based on the course artifacts that you provided, here are the assessments XXX that you mentioned, is that true?”
    2) When do you use each of these assessments?
    3) How long have you used each of these assessments in your course?
    4) Why did you decide to use them? (should go through each of them)
        a) How do you design the assessment? What resource did you use to design/choose each of these assessment?
        b) What criteria did you use when designing or writing a question for these assessments?
        c) When do you prepare assignments?
    5) What feedback each of the assessment provide to you? To what extent do you use this feedback? (do it for each assessment)
    6) What feedback each of the assessment provide to students?
    7) How does each assessment count toward students’ grades? Why are you using this grading scheme?

Appendix B. Survey Instruments

B1. DCaT Version 1

* Indicated that the item was reverse coded before analysis.

Construct	Item	Option
Cooperation	1. Instructors in my department discuss the challenges they face in the classroom with colleagues. 2. Instructors in my department share resources (e.g., idea, materials, sources, technology) about how to improve teaching with colleagues. 3. Instructors in my department use peer teaching observations to improve their teaching. 4. Instructors in my department have someone they can go to for advice about teaching.	Strongly Agree Agree Neither agree or disagree Disagree Strongly Disagree
Participation	1. How frequently do you usually participate in the decision to hire new instructional staff? 2. How frequently do you usually participate in decisions on the annual evaluation and promotion of any of the professional instructional staff? 3. How frequently do you participate in decisions on the adoption of new teaching-related policies (e.g., contact hours, program assessment)? 4. How frequently do you participate in the decisions on the adoption of new teaching methods and/or curriculum?	Very Frequently Frequently Sometimes Rarely Never
Supervisor Support	1. My department chair is willing to seek creative solutions to budgetary constraints in order to maintain adequate teaching-related support for instructors. 2. My department chair can be relied upon to give good teaching-related advice to instructors. 3. My department chair shows an understanding of the workload of instructors. 4. My department chair shows that s/he has confidence in instructors.	Strongly Agree Agree Neither agree or disagree Disagree Strongly Disagree

Warmth	1. Instructors interact informally on campus ground (e.g., coffee or lunch break). 2. Instructors within my department provides strong social support to each other (e.g., emotional, personal advice, sense of belonging, etc.). 3. Instructors readily interact socially regardless of academic rank. 4. Instructors socialize with each other on a regular basis.	Strongly Agree Agree Neither agree or disagree Disagree Strongly Disagree
Growth	1. Instructors in our department are expected to demonstrate growth in teaching skills. 2. Our department typically hosts teaching development events (e.g. talks, workshops) relevant to our discipline. 3. Instructors in our department are assigned a mentor for advice about teaching. 4. Instructors in my department are provided with resources to improve teaching (e.g., funding to travel to teaching conferences, or to purchase instructional technologies).	Strongly Agree Agree Neither agree or disagree Disagree Strongly Disagree
Innovation	1. *The departmental climate discourages instructors in our department from trying new teaching techniques. 2. Instructors in my department are “ahead of curve” when it comes to implementing innovative teaching strategies. 3. Instructors in my department are incentivized to re-design the course curriculum. 4. The departmental climate encourages instructors to be innovative in their teaching.	Strongly Agree Agree Neither agree or disagree Disagree Strongly Disagree
Autonomy	1. Instructors in my department have considerable flexibility in the content they choose to teach in their course. 2. Instructors in my department have considerable flexibility in the instructional strategies they choose to implement in their courses. 3. Instructors in my department can make their own decision without checking with anyone else about their course’s structure and organization (e.g., schedule	Strongly Agree Agree Neither agree or disagree Disagree Strongly Disagree

	of exams, office hour, grading policies, types of assignment). 4. Instructors in my department have considerable autonomy in the teaching of their courses.	
Achievement	1. My department encourages instructors to set high standards for their teaching. 2. My department encourages instructors to continuously develop their own teaching competences. 3. Instructors continuously try to outperform their own past teaching performance. 4. Instructors strive to exceed the teaching performance of fellow instructors within the department.	Strongly Agree Agree Neither agree or disagree Disagree Strongly Disagree
Outward Focus	1. *Ways of improving student learning are not given too much thought in my department. 2. *My department does not concern itself with the state and demand of the job market that undergraduate students will enter. 3. My department is continually looking for new opportunities to improve student preparedness for their careers. 4. My department pays attention to the retention of different groups of students (e.g., underrepresented group, major vs. non major) throughout the undergraduate curriculum.	Strongly Agree Agree Neither agree or disagree Disagree Strongly Disagree

Extrinsic Rewards	1. Instructors in my department are regularly nominated for campus teaching awards based on results of their teaching evaluation. 2. In my department, a teaching demonstration is required as part of the hiring process for tenure-track (or equivalent) faculty. 3. In my department, evidence of effective teaching is valued when making decisions about continued employment and/or promotion. 4. In my department, instructors with a record of teaching excellence are financially rewarded (e.g., bonuses, raises, or similar).	Strongly Agree Agree Neither agree or disagree Disagree Strongly Disagree
Performance Feedback	1. Overall instructors in my department usually receive feedback on their teaching performance. 2. *Overall instructors in my department don't have any idea how well they are teaching their courses. 3. Overall instructors in my department have difficulty measuring the quality of their teaching. 4. *The way instructors in my department teach their courses is rarely assessed.	Strongly Agree Agree Neither agree or disagree Disagree Strongly Disagree
Resources	1. Instructors in my department have flexible, physical spaces for teaching and learning. 2. Instructors in my department have adequate departmental funding to support teaching. 3. Instructors in my department have adequate time to reflect upon and make changes to their instruction. 4. Instructors in my department have access to pedagogical experts to help them to improve their teaching.	Strongly Agree Agree Neither agree or disagree Disagree Strongly Disagree

B2. DCaT Version 3

All response options for DCaT version 3 are:

Strong disagree
Disagree
Neither agree nor disagree
Agree
Strongly agree
I don't know

* Indicated that the item was reverse coded before analysis.

Construct	Item
Involvement	1. Overall instructors participate in departmental decisions regarding the process of evaluating teaching. 2. Overall instructors participate in decisions on the adoption or implementation of new, department-wide, teaching-related policies (e.g., program assessment). 3. Overall instructors in my department share resources with colleagues about how to improve teaching (e.g., ideas, materials, sources, technology). 4. Overall instructors in my department consult with each other on teaching related issues.
Supervisor Support	1. My department chair is willing to seek creative solutions to budgetary constraints in order to maintain adequate reaching-related support for instructors. 2. My department chair can be relied upon to give good teaching-related advice to instructors. 3. My department chair shows an understanding of the workload of instructors. 4. My department chair shows that s/he has confidence in instructors.
Growth	1. Overall instructors in my department are expected to demonstrate improvement in their teaching skills. 2. Overall instructors in my department are expected to participate in teaching professional development events hosted by my department or institution (e.g., workshops, talks from educational specialist). 3. Overall instructors in my department are expected to participate in regional or national teaching professional development events (e.g., teaching conferences, attending regional or national teaching workshops). 4. Overall instructors in my department are provided with enough support to improve their teaching.
Innovation	1. Overall instructors in my department are encouraged to try new teaching practices. 2. Overall instructors in my department embrace other colleagues' innovative teaching practices. 3. Overall instructors in my department are encouraged to re-design the course curriculum.

	4. Overall instructors in my department are willing to try innovative teaching practices adopted by their colleagues.
Autonomy	1. Overall instructors in my department have considerable flexibility in the content they choose to teach in their course. 2. Overall instructors in my department have considerable flexibility in the instructional strategies they choose to implement in their courses. 3. Overall instructors in my department can make their own decision without checking with anyone else about the course's structure and organizations (e.g., types/numbers of assignment). 4. Overall instructors in my department have considerable autonomy in the teaching of their courses.
Achievement	1. Overall instructors in my department set high standards for their teaching. 2. Overall instructors in my department strive to exceed the minimum standards of teaching. 3. Overall instructors in my department strive to outperform their past teaching performance. 4. Overall instructors in my department care about their teaching.
Outward Focus	1. *Ways of improving student learning are not given too much thought in my department. 2. *My department does not concern itself with the state and demand of the job market that undergraduate students will enter. 3. My department is continually looking for new opportunities to improve student preparedness for their careers. 4. My department pays attention to the retention of different groups of students (e.g., underrepresented group, major vs. non-major) throughout the undergraduate curriculum.
Performance Feedback	1. Overall instructors in my department regularly receive feedback from their colleagues on their teaching performance. 2. Overall instructors in my department receive constructive feedback about their teaching effectiveness. 3. *Overall instructors in my department have difficulties measuring the effectiveness of their teaching. 4. Teaching effectiveness in my department is mostly measured through student evaluations.

B3. DCaT Version 4 (final version)

This next section aims to measure the departmental climate with respect to teaching. We ask you to consider each question from the perspective of your department and not your college or institution.

When answering these questions, please focus on:

- The last ____ (the year they mentioned in question 6 of the demographic section)
- The undergraduate curriculum only (not the graduate curriculum).

Instructors in these questions refer to faculty who teach undergraduate level courses (including lecturers, tenured/tenure-track professors, and professors of practice but EXCLUDING graduate students)

All response options for DCaT version 4 are:

Strong disagree
Disagree
Neither agree nor disagree
Agree
Strongly agree
I don't know

* Indicated that the item was reverse coded before analysis.

Construct	Item
Involvement	1. Overall instructors participate in departmental decisions regarding the process of evaluating teaching. 2. Overall instructors participate in decisions on the adoption or implementation of new, department-wide, teaching-related policies (e.g., program assessment). 3. Overall instructors in my department share resources with colleagues about how to improve teaching (e.g., ideas, materials, sources, technology). 4. Overall instructors in my department consult with each other on teaching related issues.
Supervisor Support	1. My department chair can be relied upon to give good teaching-related advice to instructors. 2. My department chair shows an understanding of the workload of instructors. 3. My department chair shows that s/he has confidence in instructors.
Growth	1. Overall instructors in my department are expected to demonstrate improvement in their teaching skills. 2. Overall instructors in my department are expected to participate in teaching professional development events hosted by my department or institution (e.g., workshops, talks from educational specialist). 3. Overall instructors in my department are expected to participate in regional or national teaching professional development events (e.g., teaching conferences, attending regional or national teaching workshops). 4. Overall instructors in my department are provided with enough support to improve their teaching.

Innovation	1. Overall instructors in my department are encouraged to try new teaching practices. 2. Overall instructors in my department embrace other colleagues’ innovative teaching practices. 3. Overall instructors in my department are encouraged to re-design the course curriculum. 4. Overall instructors in my department are willing to try innovative teaching practices adopted by their colleagues.
Autonomy	1. Overall instructors in my department have considerable flexibility in the content they choose to teach in their course. 2. Overall instructors in my department have considerable flexibility in the instructional strategies they choose to implement in their courses. 3. Overall instructors in my department can make their own decision without checking with anyone else about the course’s structure and organizations (e.g., types/numbers of assignment). 4. Overall instructors in my department have considerable autonomy in the teaching of their courses.
Achievement	1. Overall instructors in my department set high standards for their teaching. 2. Overall instructors in my department strive to exceed the minimum standards of teaching. 3. Overall instructors in my department care about their teaching.
Outward Focus	1. *Ways of improving student learning are not given too much thought in my department. 2. *My department does not concern itself with the state and demand of the job market that undergraduate students will enter. 3. My department is continually looking for new opportunities to improve student preparedness for their careers. 4. My department pays attention to the retention of different groups of students (e.g., underrepresented group, major vs. non-major) throughout the undergraduate curriculum.
Performance Feedback	1. Overall instructors in my department regularly receive feedback from their colleagues on their teaching performance. 2. Overall instructors in my department receive constructive feedback about their teaching effectiveness. 3. *Overall instructors in my department have difficulties measuring the effectiveness of their teaching.

B4. DCaT attached Demographic questions

1. What is your academic rank?
    - Professor
    - Associate Professor
    - Assistant Professor
    - Professor of Practice
    - Associate professor of Practice
    - Assistant professor of Practice
    - Lecturer

2. What is your tenure status at your institution?
    - Tenured
    - Tenure tack, but not tenured
    - Not tenure track, but institution has tenure-system
    - Institution has no tenure system

3. Approximately what is the distribution of your appointment? The total should add to 100%. If a field is not applicable, please enter 0.

   Teaching ____
   Research ____
   Service ____
   Administration ____

4. How long have you been a professor/lecturer in your department?

5. How many different chairs have led your department since you stated your professor/lecturer position in the department?

6. How long has your current chair being in his/her current position?

7. Can you vote, at the department level, on teaching-related issues?

8. How many credit hours are you expected to teach per semester on average?

9. What type of course(s) have you taught most often in the past 1 year(s)? Choose all that apply.
    - Undergraduate course at the intro level (Freshman and sophomore)
    - Undergraduate course at the upper level (Junior and senior)
    - Graduate level courses
    - Have not taught any

10. What is/are the class sizes of the course(s) you have taught?
    - Small(0-25)
    - Medium (26-75)
    - Large(76-200)
    - Extra Large (above 201)

B5. Abbreviated Measurement Instrument for Scientific Teaching (MIST)

Please complete the survey based only on the lecture portion of that course and the last time you taught that course.

1. Indicate the average percent of class time during which students were asked to answer questions, solve problems, or complete activities <u>other than listening to a lecture (0-100%)</u>

<u>Polling Methods:</u>

Polling methods include clickers, Poll Anywhere, Learning Catalytics, colored cards, or other audience response systems that are used to determine how many students answer a question in a particular way.

2. Students were asked to use a polling method to answer questions in the classroom approximately (6-pt frequency)
   A. Zero questions
   B. 1-2 questions per month
   C. 3-4 questions per month
   D. 2-3 questions per week
   E. 4-5 questions per week
   F. 6-10 questions per week
   G. More than 10 questions per week

<u>In-class activities:</u>

In-class activities <u>other than polling questions</u> include any exercise or activity in which the students are not listening to a lecture during class, such as student discussions, worksheets, problem sets, case studies, hands-on demonstrations, role plays, concept maps, one-minute essays, think-pair-shares, inquiry-based activities, low point value quizzes, and other related activities.

3. Students were asked to complete in-class activities approximately
   A. Zero times
   B. Up to 1 activity per month
   C. 2-3 activates per month
   D. 4-7 activates per month
   E. 2-3 activates per week
   F. 4-5 activates per week
   G. More than 5 activities per week

4. Students were asked to **work in groups** of two or more on **in-class** activities, discussions, assignments, or projects other than polling questions approximately:
   A. Zero times
   B. 1-2 times during the semester
   C. About 1 time per month
   D. 2-3 times per month
   E. 1-2 times per week
   F. 3-4 times per week
   G. More than 4 times per week

<u>Out-of-class activities:</u>

Out-of-class assignments are any required exercise, activity, or project completed outside of class time, _other than_ reading a textbook or watching a video. Out-of-class assignments include worksheets, problem sets, case studies, online learning tools, inquiry-based activities, low point value quizzes, discussion boards, and other related activities.

5. Students were asked or encouraged to work in groups of two or more on **out-of-class** activities, assignments, or projects approximately
   a. Zero times
   b. 1-2 times during the semester
   c. About 1 time per month
   d. 2-3 times per month
   e. 1-2 times per week
   f. 3-4 times per week
   g. More than 4 times per week

Appendix C. Supplementary Tables and Figures

C1. Constructs measured in prior studies on organizational climate

Year	Name of survey	Author(s)	Constructs measured in the study
1963	Organizational Climate Description Questionnaire	Halpin &Croft	Disengagement, Hindrance, Esprit, Intimacy, Aloofness, Production Emphasis, Thrust in leadership, Consideration
1967	Relationship of Centralization to Other Structural Properties	Hage & Aiken	Participation, Hierarchy of authority, Job codification, Professional training, Rule observations, Professional activity
1993	Climate Survey	Ostroff	Participation, Cooperation, Warmth, Social rewards, Growth, Innovation, Autonomy, Intrinsic rewards Achievement, Hierarchy of authority, Structure, Extrinsic rewards.
1997	The Organizational Climate Questionnaire	Furnham.& Goodstein	Role clarity, Respect, Communication, Reward system, Career development, Planning and decision making, Innovation, Relationship, Teamwork and support, Quality of service, Conflict management, commitment and morale, Training and learning, Direction
2005	Organizational Climate Measure	Patterson et al	Employee welfare, Autonomy, Participation, Communication, Integration, Emphasis on training, Supervisory support, Formalization, Tradition, Flexibility, Innovation, Outward Focus, Reflectivity, Effort, Clarity of organization goal, efficiency, Quality, pressure to produce, Performance feedback
2007 **2010**	Higher Education Research Institute Faculty survey	Lee Hurtado	Collegiality, Commitment to diversity, Commitment to scholarship and scholarly recognition, Commitment to students' affective development, Commitment to teaching, Dissatisfaction with collegiate culture, Governance stress, Instrumental orientation, Job satisfaction, Multicultural orientation, Prestige orientation, Student centeredness, Valuing professional autonomy
2009	Organizational Change Questionnaire	Bouckenooghe, Devos & Van den Broeck	Trust in leadership, Politicking, Cohesion, Participation, Support by supervisors, Quality of change communication, Attitude of top management toward change, Emotional readiness for change, Cognitive readiness for change, Intentional readiness for change
2011	Department Teaching Climate	Knorek	Supportive interventions: administrative support, faculty communication, faculty development; Important practices: hiring practices, tenure/promotion; Awards
2014	Survey of Climate for Instructional Improvement	Walter, Beach, Henderson & Williams	Resources, Professional development, Collegiality, Academic freedom and autonomy, Leadership, Respect
2017	Instructional climate survey	Landrum, Viskupic, Shadle & Bullock	Free choice of teaching method, Institutional support, Teaching-research balance, Encouragement to use evidence-based instructional practices, Teacher connectedness,
2018	Collaborative on Academic Careers in Higher Education	Mamiseishvili, & Lee	Satisfaction with autonomy, Satisfaction with interactions, Satisfaction with departmental climate, Satisfaction with recognition, Global Satisfaction

C2. Raw response rates by department

Department	Discipline	School Type	Raw Response Rate
[a] Department 16	Geology	Doctoral Universities – Very high research activity	90%
[a] Department 17	Atmospheric science	Doctoral Universities – Very high research activity	77%
Department 22	Electrical and Computer Engineering	Doctoral Universities – Very high research activity	53%
Department 21	Chemistry	Master's Colleges and Universities – Smaller programs	43%
Department 6	Chemistry and Biochemistry	Doctoral Universities – High research activity	40%
Department 20	Chemistry	Doctoral/Professional Universities	39%
Department 1	Chemistry	Doctoral Universities – Very high research activity	38%
Department 8	Chemistry and Biochemistry	Doctoral Universities – Very high research activity	37%
Department 18	Chemistry	Baccalaureate Colleges	33%
Department 7	Chemistry and Biochemistry	Doctoral Universities – Very high research activity	30%
Department 14	Chemistry	Doctoral Universities – High research activity	29%
Department 10	Chemistry and Biochemistry	Doctoral Universities – Very high research activity	27%
Department 3	Chemistry & Chemical Biology	Doctoral Universities – High research activity	24%
Department 2	Chemistry	Doctoral Universities – Very high research activity	20%
Department 9	Chemistry	Doctoral Universities – Very high research activity	20%
Department 4	Chemistry and Biochemistry	Doctoral Universities – High research activity	20%
Department 5	Chemistry	Doctoral Universities – Very high research activity	17%
Department 15	Chemistry	Doctoral Universities – Very high research activity	15%
Department 12	Chemistry	Doctoral Universities – Very high research activity	13%
Department 19	Chemistry	Baccalaureate Colleges	11%
Department 13	Chemistry	Doctoral Universities – Very high research activity	10%
Department 11	Chemistry	Doctoral Universities – Very high research activity	9%

[a] *Department 16 and 17 are from the same institution.*

C3. Cleaned response rates for departments with highest raw response rates

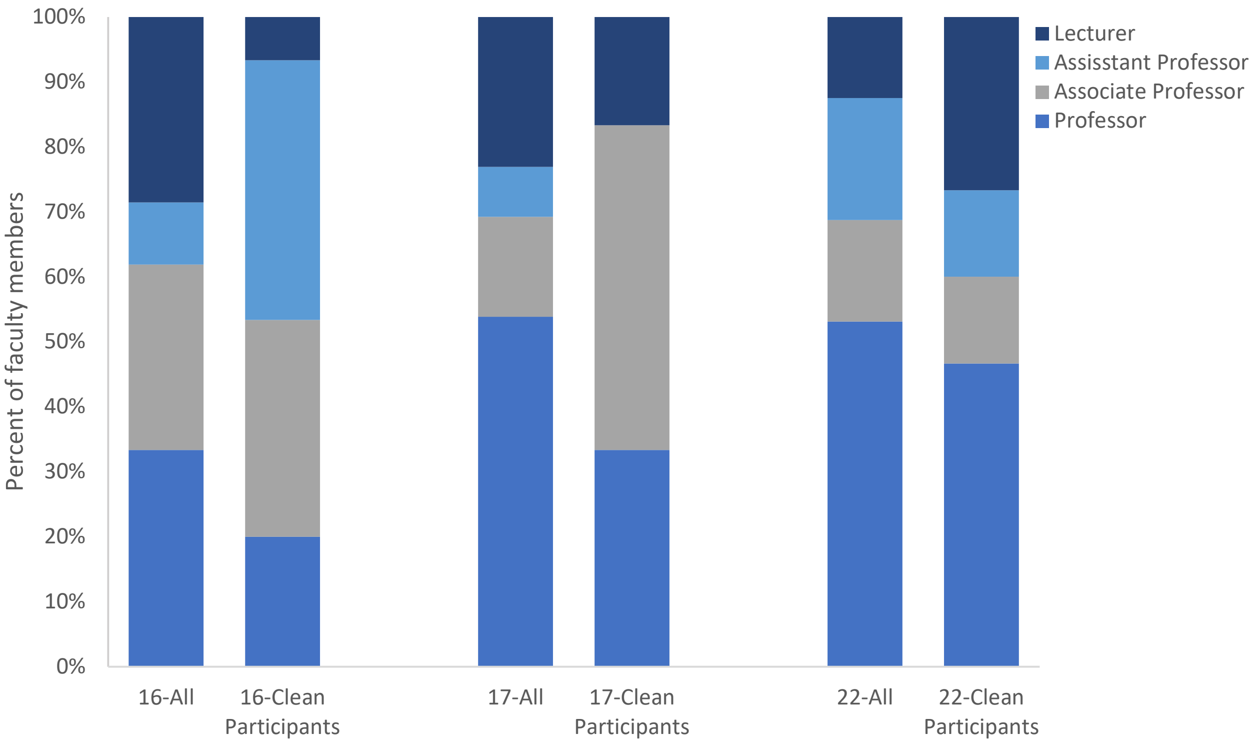
Department	Discipline	School Type	Cleaned Response Rate
[a] Department 16	Geology	Doctoral Universities – Very high research activity	71%
[a] Department 17	Atmospheric science	Doctoral Universities – Very high research activity	46%
Department 22	Electrical and Computer Engineering	Doctoral Universities – Very high research activity	44%

[a] *Department 16 and 17 are from the same institution.*

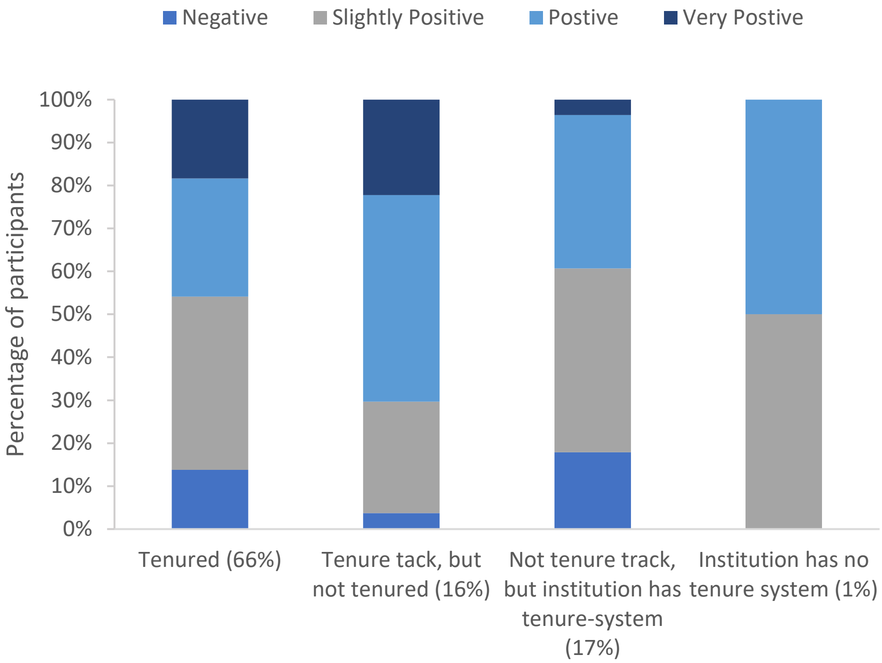
C4. Estimated model proportions and estimated means for the VEE model (n=166)

Constructs	Type of climate (Estimated model proportion)	
	1 (58%)	2 (42%)
Involvement	4.0	3.4
Growth	3.4	2.7
Autonomy	4.0	3.8
Supervisor Support	4.1	3.0
Innovation	3.6	3.1
Outward Focus	4.0	2.9
Achievement	4.2	3.6
Performance feedback	3.3	2.3

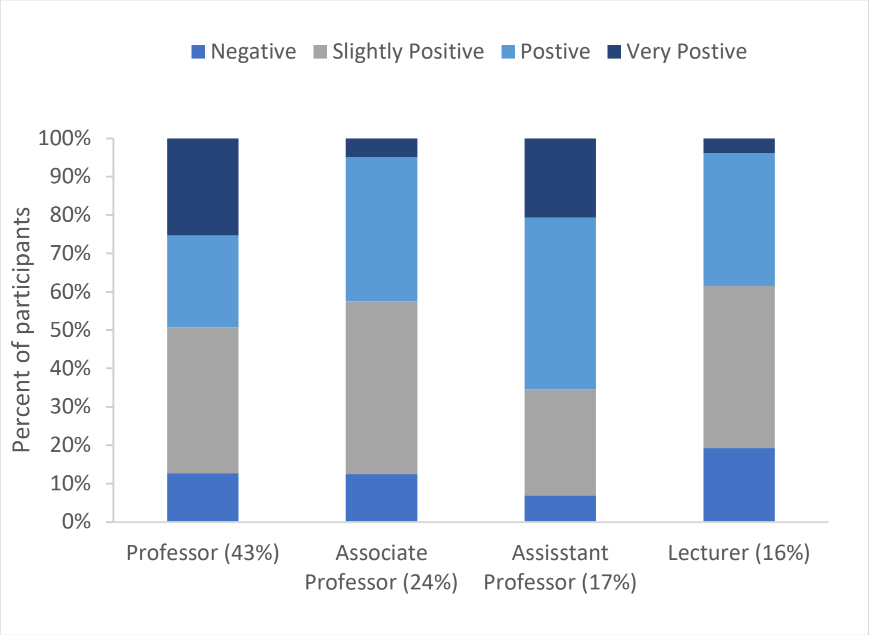
C5. Academic rank comparison of whole department versus study participants for the three departments with the highest raw response rates



C6. Tenure status versus type of psychological collective climate (N=166)



C7. Academic rank versus type of psychological collective climate (N=166)



C9. Comparison of NWA-COPUS classification and N-LCTR in “Function of Content” classification.

The alignment of NWA-COPUS classification and N-LCTR in “Function of Content” classification is highlighted in green.

NWA-COPUS Classification	The Function of Content Classification	Number of Participants	Percentage of Participants
Didactic	Instructor-Centered	4	40%
	Lower level of Transitioning	3	30%
	Higher level of Transitioning	3	30%
Lower interactive lecture	Instructor-centered	1	8%
	Lower level of Transitioning	3	25%
	Higher level of Transitioning	3	25%
	Learner-centered	5	42%
Higher interactive lecture	Instructor-centered	1	25%
	Higher level of Transitioning	1	25%
	Learner-centered	2	50%
Student-centered	Higher level of Transitioning	1	50%
	Learner-centered	1	50%

C10. Comparison of NWA-COPUS classification and N-LCTR in “The Role of Instructor” classification.

The alignment of NWA-COPUS classification and N-LCTR in “The Role of Instructor” classification is highlighted in green.

NWA-COPUS Classification	The Role of the Instructor Classification	Number of Participants	Percentage of Participants
Didactic	Instructor-centered	2	20%
	Lower level of Transitioning	4	40%
	Higher level of Transitioning	4	40%
Lower Interactive Lecture	Instructor-centered	1	8%
	Lower level of Transitioning	4	33%
	Higher level of Transitioning	5	42%
	Learner-centered	2	17%
Higher Interactive Lecture	Lower level of Transitioning	1	25%
	Higher level of Transitioning	1	25%
	Learner-centered	2	50%
Student-centered	Lower level of Transitioning	1	50%
	Learner-centered	1	50%

C11. Comparison of NWA-COPUS classification and N-LCTR in "The Responsibility for Learning" classification.

The alignment of NWA-COPUS classification and N-LCTR in "The Responsibility for Learning" classification is highlighted in green.

NWA-COPUS Classification	The Responsibility for Learning Classification	Number of Participants	Percentage of Participants
Didactic	Instructor-centered	8	80%
	Lower level of Transitioning	2	20%
Lower interactive lecture	Instructor-centered	6	50%
	Lower level of Transitioning	5	42%
	Learner-centered	1	8%
Higher interactive lecture	Lower level of Transitioning	3	75%
	Learner-centered	1	25%
Student-centered	Lower level of Transitioning	2	100%

C12. Comparison of NWA-COPUS classification and N-LCTR in "The Purpose and Processes of Assessment" classification.

The alignment of NWA-COPUS classification and N-LCTR in "The Purpose and Processes of Assessment" classification is highlighted in green.

NWA-COPUS Classification	The Purpose and Processes of Student Assessment Classification	Number of Participants	Percentage of Participants
Didactic	Instructor-centered	7	70%
	Lower level of transitioning	3	30%
Lower interactive lecture	Instructor-centered	2	17%
	Lower level of transitioning	8	67%
	Higher level of transitioning	2	17%
Higher interactive lecture	Lower level of transitioning	3	75%
	Learner-centered	1	25%
Student-centered	Instructor-centered	1	50%
	Lower level of transitioning	1	50%

C13. Comparison of NWA-COPUS classification and N-LCTR in “The Balance of Power” classification.
The alignment of NWA-COPUS classification and N-LCTR in “The Balance of Power” classification is highlighted in green.

NWA-COPUS Classification	The Balance of Power Classification	Number of Participants	Percentage of Participants
Didactic	Instructor-centered	4	40%
	Lower level of Transitioning	4	40%
	Higher level of Transitioning	2	20%
Lower interactive lecture	Instructor-centered	3	25%
	Lower level of Transitioning	3	25%
	Higher level of Transitioning	6	50%
Higher interactive lecture	Instructor-centered	1	25%
	Lower level of Transitioning	3	75%
Student-centered	Higher level of Transitioning	2	100%

C14. Comparison of N-LCTR classification and NWA-COPUS classification by sampling intensity.
Alignment between the two classifications is highlighted in green.

Number of Classroom Clips	NWA-COPUS Classification	N-LCTR classification	Number of Participants
≥6	Didactic	Instructor-centered	1
	Lower interactive lecture	Lower level of Transitioning	2
5	Didactic	Lower level of Transitioning	1
	Lower interactive lecture	Lower level of Transitioning	1
	Lower interactive lecture	Higher level of Transitioning	1
	Higher interactive lecture	Lower level of Transitioning	1
4	Didactic	Instructor-centered	1
	Didactic	Lower level of Transitioning	1
	Lower interactive lecture	Lower level of Transitioning	3
	Lower interactive lecture	Higher level of Transitioning	2
	Higher interactive lecture	Lower level of Transitioning	1
	Higher interactive lecture	Higher level of Transitioning	1
3	Didactic	Instructor-centered	1
	Didactic	Lower level of Transitioning	1
	Lower interactive lecture	Lower level of Transitioning	2
	Learner-centered	Higher level of Transitioning	2
2	Didactic	Instructor-centered	1
	Didactic	Lower level of Transitioning	3
	Lower interactive lecture	Lower level of Transitioning	1
	Higher interactive lecture	Lower level of Transitioning	1

C15. Instructors' rationale of utilizing assessment

Rationale for utilizing assessment	Definition	Participants, n
Opportunity for instructors to evaluate student learning	assessment is identified as way for instructor to evaluate what knowledge and skills that the students have learned from the course, and where students at the stage of their learning	16
Probe students' preparation for next course or major	assessment is used as a way to measure if students are ready for their next course (e.g., upper-level chemistry course (Organic chem, Physical chem), or other courses that required by their major (e.g., biology, engineering)	12
Opportunity for instructor to get feedback from students and modify their instruction	assessment is an opportunity for instructor to reflect on their teaching practices, and identify ways to better help students for their learning	6
Assign grades	Assessment is a way to provide points for students	6
Tradition	assessment is used as tradition to measure students' learning which is what the instructor experienced as a student, or they learned when they start teaching, or it is what other instructors did in their home institution/department	5
Opportunity for students to get feedback on their understanding	assessment is identified as a way to for instructor to provide feedback on students' learning progress.	5
To ensure students are studying and learning	assessment is identified as a way to ensure students are keeping learning the material along the semester and keep up with the progress of the course	4
Opportunity for students to evaluate their learning	assessment is identified as an opportunity for students to reflect on what they have learned from the course, what they still don't know about the course and think about metacognition	4
Opportunity for students to demonstrate understanding	assessment is identified as an opportunity for students to demonstrate their understanding to instructor about what they have learned in the course	4
Probe students' ability to solve problems in new scenario	assessment used as a way to probe if students can apply what they learned in a new scenario	2

C16. Instructors' assessment practices

Code	**definition**
Homework	Assignment that students did after the class through different online platforms or through paper-pencil
Cumulative Final	An exam given at the end of the semester to evaluate students' learning of ALL the material has been taught during the course
Mid-terms	Assessment that happened during the semester to measure students' understanding of certain amount of the course material
Quiz	Includes daily quizzes at the beginning of each class, quizzes for each unit-specification grading, weekly quizzes, knowledge check every couple of weeks
Poll Questions	Assessment that instructor used during the class time to check students' understanding through clickers, poll everywhere, plickers or other polling methods
Pre-class Assignments	Assessment that students need to finish before they come to the class, e.g., readings related to some fundamental concepts, videos to introduce fundamental concepts
In-class Worksheet	Assessment assigned during the in-class time, like PLTL, worksheet day, team learning (e.g., Chemistry connect, which ask students to work as a group to talk about chemistry in the real world) which involves doing worksheet
In-class Questioning	Students actively participated during the class by doing group discussion, Q and A, informal group discussion- not graded activities
Reflective Assessment	Assessment that helps students to reflect on their own learning process, e.g., Muddiest point at the end of a class; exam/ homework correction; student' self-surveys used to assess their learning gains
Practice Assessment	Exam prep or practice quizzes which used to help students prepare for their exam
Group Project	An assignment which needs a group of students to work together as a project
Pre-course Assignments	Assignment that students need to finish before the course which helps students to be prepared for their math skills or some basic high school chemistry
Group Exams	Exam students need to finish as a group, and it is graded
Other Final	No cumulative final, the cumulative final was let students to attempt on any learning objectives that they might still be missing. (Specification grading)

C17. Instructors' rationale in utilizing a specific type of assessment

Homework (n=19)

Code	***Number of Participants***
opportunity for students to learn and practice skills	7
N/A	7
opportunity for students to get feedback on their understanding	4
opportunity to review material	4
prepare students for next higher stakes assessment(probe students' preparation for next level of assessment)	2
Boost grades	2
to ensure students are studying and learning	2
Tradition	2
opportunity for instructors to evaluate student learning	1
provide points	1
encourage students to relearn things they did not understand	*1*
promote independent learning	1
probe students' engagement with the content	1
use varied types of assessment to promote students' learning	1
opportunity for students to demonstrate their retention of knowledge	1

Mid-Term (n= 18)

Code	***Number of Participants***
Tradition	9
opportunity for students to demonstrate understanding	2
opportunity to encourage students to use learning resources	2
prepare students for next higher stakes assessment	2
relieve stress- the whole section may belong to "assessment development criteria".	2
opportunity for practice skills or conceptual understanding	2
N/A	2
opportunity to review material	1
opportunity for students to demonstrate their retention of knowledge	1
relieve stress	1
probe individual understanding	1
opportunity for students to evaluate their learning	1
opportunity for instructors to evaluate student learning	1
opportunity for students to get feedback on their understanding	1
to get feedback to the instructor on students understanding of the material and address practice accordingly	1
probe individual understanding	1
relieve stress	1
to ensure students are studying and learning	1
previous experience	1
to probe student time managment skills and anxiety level	1
provide points	1

Cumulative Final (n= 18)

Code	***Number of Participants***
tradition	9
N/A	6
probe students' preparation for next course or major	2
to probe whether students have learned the knowledge presented	1
to probe whether students across sections of the course are performing similarly	1
Probe students' ability to solve problems in new scenario	1
to ensure students are studying and learning	1
previous experience	1
to probe student time management skills and anxiety level	1
provide points	1

Quiz (n=13)

Code	***Number of Participants***
N/A	5
prepare students for next higher stakes assessment(probe students' preparation for next level of assessment)	4
to ensure students are studying and learning	3
opportunity to engage students	2
previous experience	2
opportunity for students to demonstrate understanding	1
opportunity for discussion between students	1
opportunity to encourage students to use learning resources	1
encourage students to relearn things they did not understand	1
frequent assessment promote engagement with materials	1
opportunity to engage students	1

Poll Questions (n=10)

Code	***Number of Participants***
N/A	5
opportunity for discussion between students	3
Opportunity for instructor to get feedback from students and modify their instruction	2
opportunity to engage students	1
opportunity for students to evaluate their learning	1
relieve stress	1

Pre-Class Assignment (n=9)

Code	***Number of Participants***
N/A	7
opportunity to review material	1
opportunity for practice skills or conceptual understanding	1
opportunity for students to demonstrate understanding	1

Class Participation (n=7)

Code	***Number of Participants***
N/A	6
Boost grades	1

In-class Worksheet (n=6)

Code	***Number of Participants***
N/A	3
to ensure students are studying and learning	1
opportunity for discussion between students	1
opportunity to encourage students to use learning resources	1
opportunity for students to get feedback on their understanding	1
Groupwork Activities (n=6)	
Code	***Number of Participants***
N/A	4
opportunity for discussion between students	1
opportunity to engage students	1
Reflective Assessment (n=3)	
Code	***Number of Participants***
N/A	3
Practice Assessment (n=2)	
Code	***Number of Participants***
N/A	3
Group Project (n=2)	
Code	***Number of Participants***
opportunity for students to demonstrate understanding	1
opportunity for discussion between students	1
N/A	1

Appendix D. Modified LCTR Rubrics

Rubric	Components
I.The Function of Content	1. In addition to building a knowledge base, the instructor uses content to help students understand why they need to learn the content for courses within or outside their major and for their future career. 2. The instructor uses content to practice using scientific practices in the discipline or to solve authentic problems. 3. The instructor helps students acquire in-depth conceptual understanding of the content to facilitate deep learning and use of transferable skills.
II.The Role of The Instructor	1. The instructor creates a supportive and success-oriented environment for learning and for accomplishment for all students through proactive, clear, and overt course-specific techniques. 2. The instructor uses diverse teaching strategies that promote the achievement of student learning. 3. The instructor develops and uses a variety of learning outcomes/goals. 4. The instructor aligns two essential components of a course: learning outcomes/goals and assessment methods in terms of content and uses consistent verbs representing the same cognitive processing demands/intellectual skills placed on the students.
III.The Responsibility for Learning	1. The instructor sets student expectations, which enable the responsibility for learning to be shared between the instructor and the students. 2. The instructor fosters students' engagement in reflection and critical review of their learning through well-structured activities.
IV.The Purposes and Processes of Student Assessment	1. The instructor uses formative assessment within the learning process. 2. The instructor promotes students to use peer-assessments. 3. The instructor allows the students to demonstrate mastery of the objectives and ability to learn from mistakes. 4. The instructor uses authentic assessments (e.g., research report, case study).
V.The Balance of Power	1. The instructor allows for some flexibility of course policies, assessment methods, learning methods, and deadlines or how students earn grades.

Appendix E. Descriptions of NWA-COPUS classifications

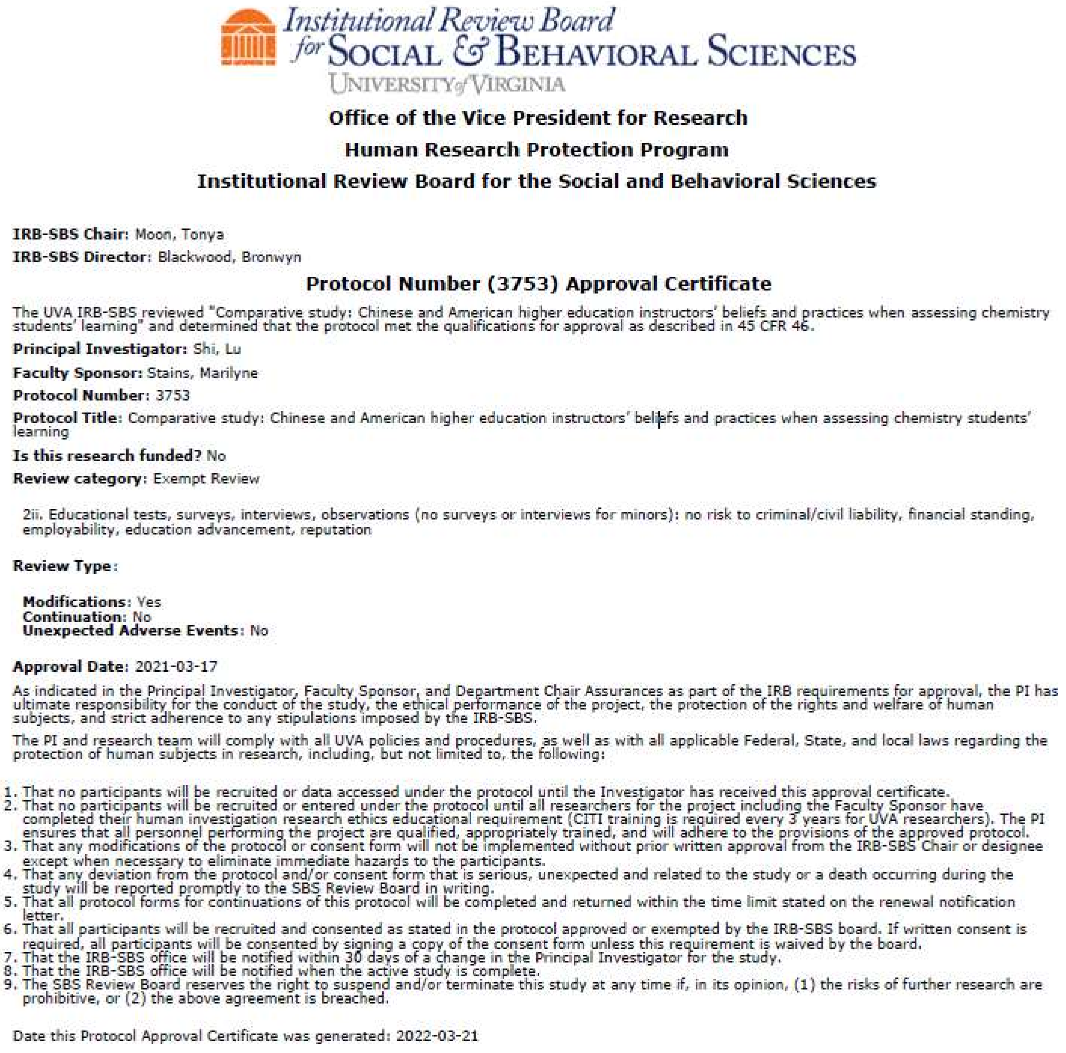
Didactic **(0-0.25)**	**Lower Interactive Lecture** **(0.26-0.50)**	**Higher Interactive Lecture** **(0.51-0.75)**	**Student-centered** **(0.76-1.00)**
At least 50% of the observed classes were classified as didactic (e.g., on average, 86% of the class time is spent lecturing)	At least 75% of the observed classes were classified as interactive lecture (e.g., on average, 74% of the class time is spent lecturing,) OR 50% or less of the observed classes were classified as student-centered (e.g., on average 52% of the class time is spent lecturing)	50% of the observed classes were classified as interactive lecture (e.g., on average, 74% of the class time is spent lecturing) and 50% as student-centered (e.g., on average, less than 53% of the class time is spent lecturing,)	At least 67% of the observed classes were classified as student-centered (e.g., on average, less than 53% of the class time is spent lecturing)

Appendix F. Description of N-LCTR groups by each LCTR rubric

	LCTR rubrics				
N-LCTR Classification	The Function of Content: How the students will use and why they should learn the content	The Role of the Instructor: Instructor should facilitate and create an environment to promote students' learning	The Responsibility for Learning: Instructors can teach students to take responsibility for their learning	The Purposes and Processes of Assessment: Assessment should be given during the learning process	The Balance of Power: Students have more control over their learning
Instructor-centered (0-0.25)	The instructor 1) only helps students to build a knowledge base, but does not use content to help students understand why they need to learn the content; 2) provides no opportunities for student to practice using scientific practice or to solve authentic problems; 3) does not provide opportunities for students to apply knowledge to new content or encourage the development of transferable skills.	The instructor 1) does not make themselves/resources available beyond office hours; 2) primarily lectures; 3) does not develop or not use (or have vague) learning outcomes/goals; 4) does not align learning outcomes/goals with assessment methods.	The instructor 1) takes the responsibility for student learning; 2) provides no opportunities for student reflection, critical review, and self-assessment of their learning.	The instructor 1) does not integrate formative assessment within the learning process; 2) does not provide students with opportunities for peer-assessment; 3) does not provide any opportunities for students (or only give opportunities to some students) to demonstrate that they learned from mistakes or to show mastery beyond the first attempt; 4) rarely or never uses authentic assessment.	The instructor mandates all policies, learning and assessment methods and deadline, or allows some students flexibility in deadlines or how they earn grades but does not let other students know about these possibilities.
Lower level of transitioning (0.26-0.50)	The instructor 1) helps student to recognize why they need to learn the content; 2) provides limited opportunities for students to practice using scientific practices or to solve authentic problems but does not teach how to do it or only teaches if students ask in one-on-one interaction; 3) provides limited opportunities to apply knowledge to new content and to develop transferable skills.	The instructor 1) is available to help students in one-on-one situations; 2) uses limited number of teaching strategies other than lecturing; 3) develops low-level learning outcomes/goals only (based on Bloom's taxonomy), and may share them with students; 4) aligns learning outcomes/goals and assessment methods but does not explain this to students.	The instructor 1) believes that students and instructors should share responsibility for learning; 2) provides opportunities for student reflection, critical review, and self-assessment of their learning, but does not teach students how to do self-assessment.	The instructor 1) minimally integrates formative assessment within the learning process; 2) requires students to use peer-assessments once or twice, and does not count this toward the final grade; 3) provides one opportunity (e.g., cumulative final) for students to demonstrate that they mastered the material after the first attempt; 4) uses one authentic assessment that is weighted 10-20 % of the final grade.	The instructor is flexible on one of the following: course policies, learning or assessment methods, deadlines, how students earn grades, and allows all students the same flexibility.

Higher level of transitioning (0.51-0.75)	The instructor 1) provides several opportunities to help students identify why they need to learn the content for use in courses within or outside their major and for their careers; 2) provides students several opportunities to practice scientific practices or to solve authentic problems and teaches this to all student in the class; 3) requires all students to use content in other contexts.	The instructor 1) in one-on-one situations gives clear and overt specific techniques to help students to succeed; 2) uses several teaching strategies other than lecturing; 3) develops low- and high-level learning outcomes/goals, and shares them with students in the syllabus; 4) aligns learning outcomes/goals and assessment methods, and shares them with students in course documents, but does not explain to students why it is important.	The instructor 1) believes that students should take responsibility for their learning; 2) provides opportunities throughout the course for student reflection, critical review, and self-assessment of their learning, and intentionally teaches students how to do self-assessment.	The instructor 1) somewhat integrates formative assessment within the learning process; 2) teaches students how to meaningfully conduct peer-assessments, uses it several times during a course, and it counts towards the final grade; 3) provides several opportunities for students to demonstrate that they mastered the material; 4) uses one authentic assessment (or assessment have authentic elements) that is weighted 21-50 % of the final grade.	The instructor is flexible on at least two of the following: course policies, learning or assessment methods, deadlines, and how students earn grades course policies, and allows students to provide feedback on policies, methods and deadlines.
Learner-centered (0.76-1.00)	The instructor 1) periodically (at least weekly) requires students to reflect or evaluate in majority of their assessment why they need to learn content for use in courses within or outside their major and for their careers; 2) provides (at least weekly) student practice to develop scientific practices or to solve authentic problems and assesses them; 3) requires students to learn or integrate content on their own in a large paper or project. This assignment counts towards their final grade.	The instructor 1) gives clear and overt specific techniques to all students and explain to students how these methods promote students success; 2) uses many teaching strategies and regularly uses active learning approaches; 3) places in the syllabus low- and high-level learning outcomes/goals and regularly refers to them throughout the course; 4) aligns learning outcomes/goals and assessment methods. This is clearly shown and explained and is actually used to align course materials.	The instructor 1) encourages students to take responsibility for learning and provides opportunities and teaches them how to do it; 2) provides formative feedback and assess students' self-assessment which includes facilitates students to develop reflection, critical review, and self-assessment skills of their learning, and explains how and why they need to do self-assessment.	The instructor 1) mostly integrates formative assessment within the learning process; 2) models how to conduct meaningful peer-assessments, uses it several time in the course, provides feedback to students, and counts it toward the final grade; 3) provides many opportunities for students to demonstrate what they mastered; 4) uses authentic assessments throughout the course. These authentic assessment elements count toward 51-80 % of the final grade.	The instructor is flexible on at least four of the following: course policies, learning or assessment methods, deadlines, and how students earn grades course policies, and actively seeks student feedback on policies, methods and deadlines and responds to their feedback with appropriate and considered changes.

Appendix G. Approved Institutional Review Board (IRB) letters



Institutional Review Board for Social & Behavioral Sciences
University of Virginia

Office of the Vice President for Research

Human Research Protection Program

Institutional Review Board for the Social and Behavioral Sciences

IRB-SBS Chair: Moon, Tonya

IRB-SBS Director: Blackwood, Bronwyn

Protocol Number (3753) Approval Certificate

The UVA IRB-SBS reviewed "Comparative study: Chinese and American higher education instructors' beliefs and practices when assessing chemistry students' learning" and determined that the protocol met the qualifications for approval as described in 45 CFR 46.

Principal Investigator: Shi, Lu

Faculty Sponsor: Stains, Marilyne

Protocol Number: 3753

Protocol Title: Comparative study: Chinese and American higher education instructors' beliefs and practices when assessing chemistry students' learning

Is this research funded? No

Review category: Exempt Review

2ii. Educational tests, surveys, interviews, observations (no surveys or interviews for minors): no risk to criminal/civil liability, financial standing, employability, education advancement, reputation

Review Type:

Modifications: Yes
Continuation: No
Unexpected Adverse Events: No

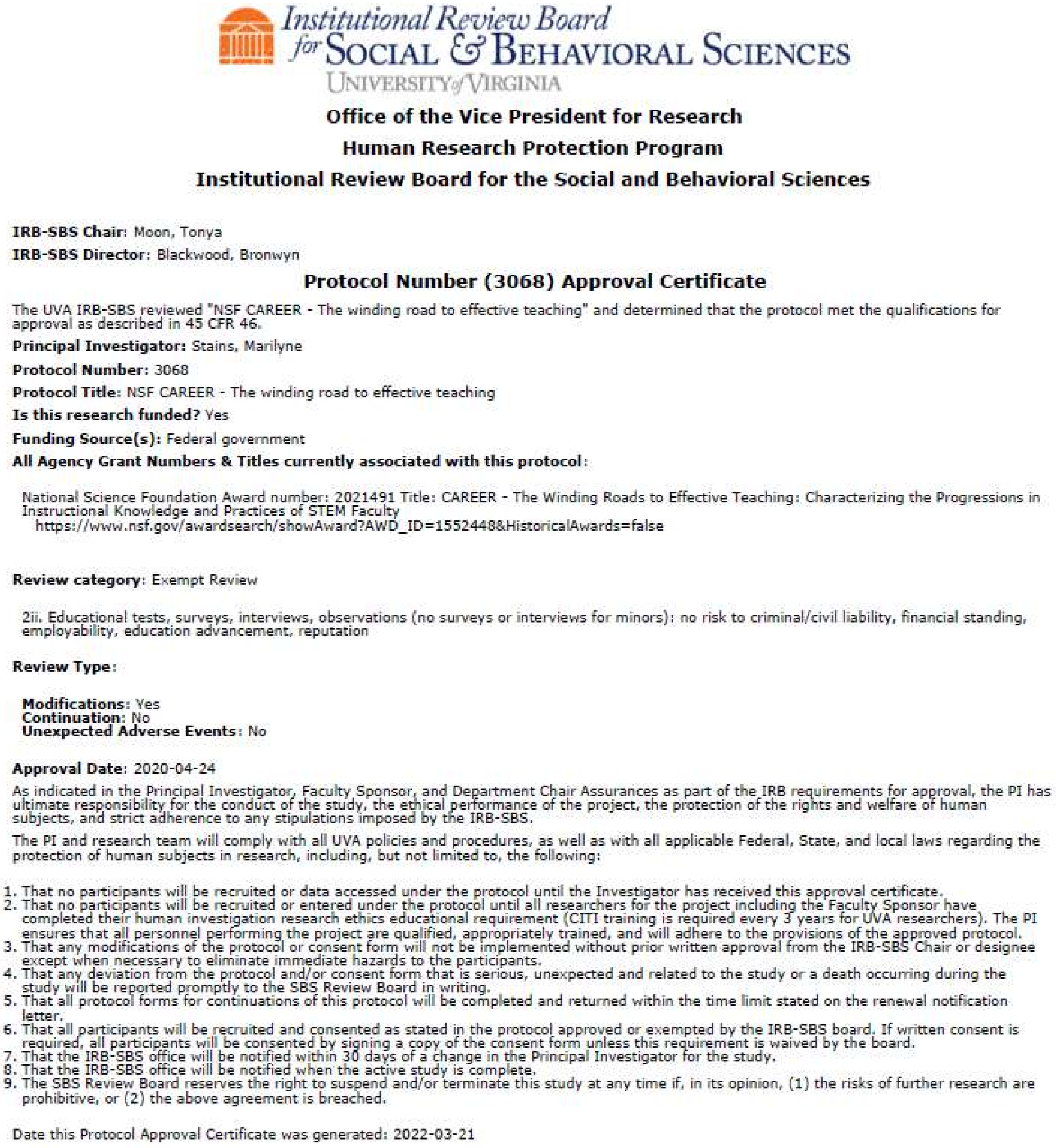
Approval Date: 2021-03-17

As indicated in the Principal Investigator, Faculty Sponsor, and Department Chair Assurances as part of the IRB requirements for approval, the PI has ultimate responsibility for the conduct of the study, the ethical performance of the project, the protection of the rights and welfare of human subjects, and strict adherence to any stipulations imposed by the IRB-SBS.

The PI and research team will comply with all UVA policies and procedures, as well as with all applicable Federal, State, and local laws regarding the protection of human subjects in research, including, but not limited to, the following:

1. That no participants will be recruited or data accessed under the protocol until the Investigator has received this approval certificate.
2. That no participants will be recruited or entered under the protocol until all researchers for the project including the Faculty Sponsor have completed their human investigation research ethics educational requirement (CITI training is required every 3 years for UVA researchers). The PI ensures that all personnel performing the project are qualified, appropriately trained, and will adhere to the provisions of the approved protocol.
3. That any modifications of the protocol or consent form will not be implemented without prior written approval from the IRB-SBS Chair or designee except when necessary to eliminate immediate hazards to the participants.
4. That any deviation from the protocol and/or consent form that is serious, unexpected and related to the study or a death occurring during the study will be reported promptly to the SBS Review Board in writing.
5. That all protocol forms for continuations of this protocol will be completed and returned within the time limit stated on the renewal notification letter.
6. That all participants will be recruited and consented as stated in the protocol approved or exempted by the IRB-SBS board. If written consent is required, all participants will be consented by signing a copy of the consent form unless this requirement is waived by the board.
7. That the IRB-SBS office will be notified within 30 days of a change in the Principal Investigator for the study.
8. That the IRB-SBS office will be notified when the active study is complete.
9. The SBS Review Board reserves the right to suspend and/or terminate this study at any time if, in its opinion, (1) the risks of further research are prohibitive, or (2) the above agreement is breached.

Date this Protocol Approval Certificate was generated: 2022-03-21



Institutional Review Board for Social & Behavioral Sciences
University of Virginia

Office of the Vice President for Research

Human Research Protection Program

Institutional Review Board for the Social and Behavioral Sciences

IRB-SBS Chair: Moon, Tonya

IRB-SBS Director: Blackwood, Bronwyn

Protocol Number (3068) Approval Certificate

The UVA IRB-SBS reviewed "NSF CAREER - The winding road to effective teaching" and determined that the protocol met the qualifications for approval as described in 45 CFR 46.

Principal Investigator: Stains, Marilyne

Protocol Number: 3068

Protocol Title: NSF CAREER - The winding road to effective teaching

Is this research funded? Yes

Funding Source(s): Federal government

All Agency Grant Numbers & Titles currently associated with this protocol:

National Science Foundation Award number: 2021491 Title: CAREER - The Winding Roads to Effective Teaching: Characterizing the Progressions in Instructional Knowledge and Practices of STEM Faculty
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1552448&HistoricalAwards=false

Review category: Exempt Review

2ii. Educational tests, surveys, interviews, observations (no surveys or interviews for minors): no risk to criminal/civil liability, financial standing, employability, education advancement, reputation

Review Type:

Modifications: Yes
Continuation: No
Unexpected Adverse Events: No

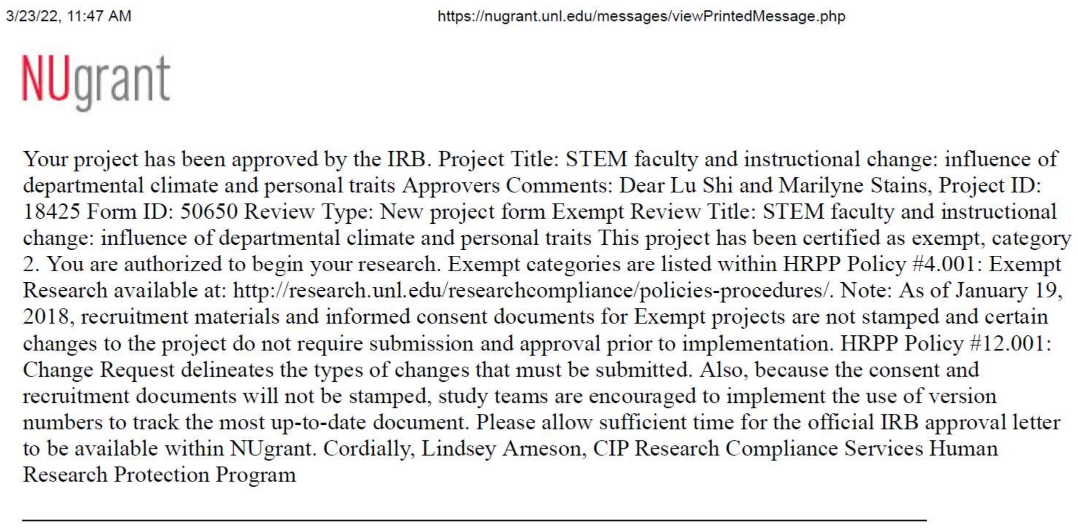
Approval Date: 2020-04-24

As indicated in the Principal Investigator, Faculty Sponsor, and Department Chair Assurances as part of the IRB requirements for approval, the PI has ultimate responsibility for the conduct of the study, the ethical performance of the project, the protection of the rights and welfare of human subjects, and strict adherence to any stipulations imposed by the IRB-SBS.

The PI and research team will comply with all UVA policies and procedures, as well as with all applicable Federal, State, and local laws regarding the protection of human subjects in research, including, but not limited to, the following:

1. That no participants will be recruited or data accessed under the protocol until the Investigator has received this approval certificate.
2. That no participants will be recruited or entered under the protocol until all researchers for the project including the Faculty Sponsor have completed their human investigation research ethics educational requirement (CITI training is required every 3 years for UVA researchers). The PI ensures that all personnel performing the project are qualified, appropriately trained, and will adhere to the provisions of the approved protocol.
3. That any modifications of the protocol or consent form will not be implemented without prior written approval from the IRB-SBS Chair or designee except when necessary to eliminate immediate hazards to the participants.
4. That any deviation from the protocol and/or consent form that is serious, unexpected and related to the study or a death occurring during the study will be reported promptly to the SBS Review Board in writing.
5. That all protocol forms for continuations of this protocol will be completed and returned within the time limit stated on the renewal notification letter.
6. That all participants will be recruited and consented as stated in the protocol approved or exempted by the IRB-SBS board. If written consent is required, all participants will be consented by signing a copy of the consent form unless this requirement is waived by the board.
7. That the IRB-SBS office will be notified within 30 days of a change in the Principal Investigator for the study.
8. That the IRB-SBS office will be notified when the active study is complete.
9. The SBS Review Board reserves the right to suspend and/or terminate this study at any time if, in its opinion, (1) the risks of further research are prohibitive, or (2) the above agreement is breached.

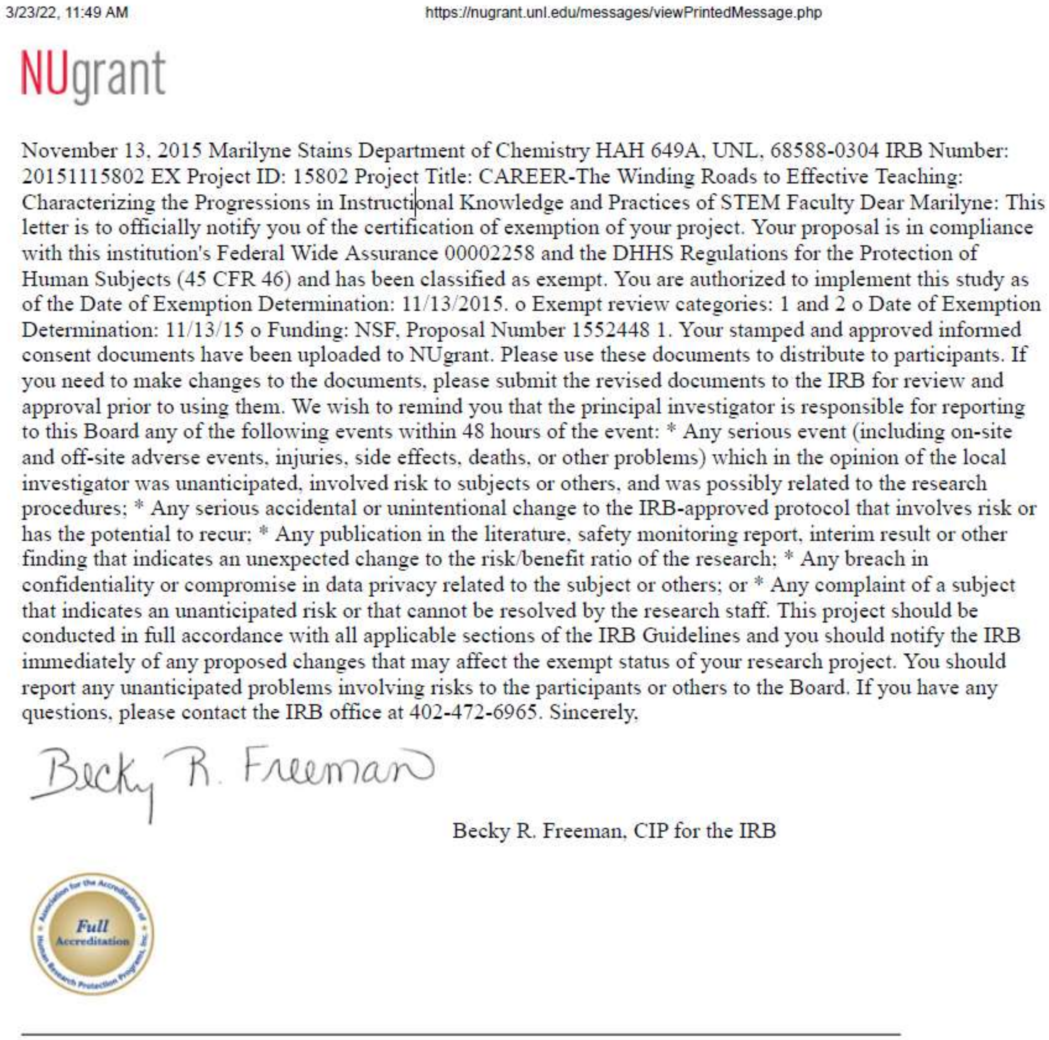
Date this Protocol Approval Certificate was generated: 2022-03-21



3/23/22, 11:47 AM https://nugrant.unl.edu/messages/viewPrintedMessage.php

NUgrant

Your project has been approved by the IRB. Project Title: STEM faculty and instructional change: influence of departmental climate and personal traits Approvers Comments: Dear Lu Shi and Marilyne Stains, Project ID: 18425 Form ID: 50650 Review Type: New project form Exempt Review Title: STEM faculty and instructional change: influence of departmental climate and personal traits This project has been certified as exempt, category 2. You are authorized to begin your research. Exempt categories are listed within HRPP Policy #4.001: Exempt Research available at: http://research.unl.edu/researchcompliance/policies-procedures/. Note: As of January 19, 2018, recruitment materials and informed consent documents for Exempt projects are not stamped and certain changes to the project do not require submission and approval prior to implementation. HRPP Policy #12.001: Change Request delineates the types of changes that must be submitted. Also, because the consent and recruitment documents will not be stamped, study teams are encouraged to implement the use of version numbers to track the most up-to-date document. Please allow sufficient time for the official IRB approval letter to be available within NUgrant. Cordially, Lindsey Arneson, CIP Research Compliance Services Human Research Protection Program



3/23/22, 11:49 AM https://nugrant.unl.edu/messages/viewPrintedMessage.php

NUgrant

November 13, 2015 Marilyne Stains Department of Chemistry HAH 649A, UNL, 68588-0304 IRB Number: 20151115802 EX Project ID: 15802 Project Title: CAREER-The Winding Roads to Effective Teaching: Characterizing the Progressions in Instructional Knowledge and Practices of STEM Faculty Dear Marilyne: This letter is to officially notify you of the certification of exemption of your project. Your proposal is in compliance with this institution's Federal Wide Assurance 00002258 and the DHHS Regulations for the Protection of Human Subjects (45 CFR 46) and has been classified as exempt. You are authorized to implement this study as of the Date of Exemption Determination: 11/13/2015. o Exempt review categories: 1 and 2 o Date of Exemption Determination: 11/13/15 o Funding: NSF, Proposal Number 1552448 1. Your stamped and approved informed consent documents have been uploaded to NUgrant. Please use these documents to distribute to participants. If you need to make changes to the documents, please submit the revised documents to the IRB for review and approval prior to using them. We wish to remind you that the principal investigator is responsible for reporting to this Board any of the following events within 48 hours of the event: * Any serious event (including on-site and off-site adverse events, injuries, side effects, deaths, or other problems) which in the opinion of the local investigator was unanticipated, involved risk to subjects or others, and was possibly related to the research procedures; * Any serious accidental or unintentional change to the IRB-approved protocol that involves risk or has the potential to recur; * Any publication in the literature, safety monitoring report, interim result or other finding that indicates an unexpected change to the risk/benefit ratio of the research; * Any breach in confidentiality or compromise in data privacy related to the subject or others; or * Any complaint of a subject that indicates an unanticipated risk or that cannot be resolved by the research staff. This project should be conducted in full accordance with all applicable sections of the IRB Guidelines and you should notify the IRB immediately of any proposed changes that may affect the exempt status of your research project. You should report any unanticipated problems involving risks to the participants or others to the Board. If you have any questions, please contact the IRB office at 402-472-6965. Sincerely,

Becky R. Freeman

Becky R. Freeman, CIP for the IRB

Association for the Accreditation of Human Research Protection Programs, Inc.
Full Accreditation