Statistics Graduate Students'
Professional Development for Teaching:
A Communities of Practice Model

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Nicola Justice

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Joan Garfield, Adviser
Andrew Zieffler, Co-Adviser

May 2017

Acknowledgements

As with many big endeavors, it took a team effort to complete this dissertation.

Huge thanks to my husband, Michael. Thank you for encouraging me when I was ready to give up. Thank you for reflecting God's love to me so brightly, beautifully, unconditionally, and resolutely. You've helped me believe God could love me deeper than I ever imagined.

Thanks to Joan, Andy, Bob, who taught me so much about research and writing, asked tough questions, gave such wonderful feedback, offered tissues for many tears, and who established the University of Minnesota's Statistics Education doctoral program that was such an amazing opportunity for me and other students.

Thanks to my family. To my "old" family (Mom, Dad, Michelle & family) — thank you for being ever-supportive, encouraging, affirming as I pursued my passions. To my "new" family (Debbie, Dick, Tiff, Nicole, & families) — thanks for being supportive of our time in Minnesota and encouraging us even though it was at times tough that we were so far away.

Thanks to my dear friend Anelise, especially for your prayers. To Anelise and all the other members of the UMN Stat Ed family who shared ideas, resources, joys, and sorrows: Liz, Ethan, Laura, Matt, Laura, Jonathan, Mike, and Rebekah (wherever you are we still love you), Suzanne, and Michelle — you are such a wonderful crew and it is a pleasure to share lives and share this discipline with you.

Thanks also to new friends in Minnesota who helped us feel at home here: the Olson family, the Hohenstein family, Liz & Scott, Brian & Stefano, Corrie & Josh, Sonja & Dave. Thanks to Hope Community Church for providing a gracious church home for us to hear the gospel. Thanks to old friends back home who have been there for us even though we haven't been there, particularly: Beth, Heather, Sienna, Hilary, Karl, Jill.

And thanks to God, for providing passions in my heart and giving me the strength and abilities to pursue them. Thanks for giving me hope and a future that could not be shaken among all the difficulties that came along the way.

Dedication

"For from Him and through Him and for Him are all things.
To Him be the glory forever!"
Romans 11:36

Abstract

Graduate teaching assistants (GTAs) are responsible for instructing approximately 25% of introductory statistics courses in the United States (Blair, Kirkman, & Maxwell, 2013). Most research on GTA professional development focuses on structured activities (e.g., courses, workshops) that have been developed to improve GTAs' pedagogy and content knowledge. Few studies take into account the social contexts of GTAs' professional development. However, GTAs perceive their social interactions with other GTAs to be a vital part of their preparation and support for teaching (e.g., Staton & Darling, 1989).

Communities of practice (CoPs) are one way to bring together the study of the social contexts and structured activities of GTA professional development. CoPs are defined as groups of practitioners who deepen their knowledge and expertise by interacting with each other on an ongoing basis (e.g., Lave & Wenger, 1991). Graduate students may participate in CoPs related to teaching in many ways, including attending courses or workshops, participating in weekly meetings, engaging in informal discussions about teaching, or participating in e-mail conversations related to teaching tasks.

This study explored the relationship between statistics graduate students' experiences in CoPs and the extent to which they hold student-centered teaching beliefs. A framework for characterizing GTAs' experiences in CoPs was described and a theoretical model relating these characteristics to GTAs' beliefs was developed. To gather data to test the model, the *Graduate Students' Experiences Teaching Statistics* (GETS) *Inventory* was created. Items were written to collect information about GTAs'

current teaching beliefs, teaching beliefs before entering their degree programs, characteristics of GTAs' experiences in CoPs, and demographic information. Using an online program, the GETS Inventory was administered to N=218 statistics graduate students representing 37 institutions in 24 different U.S. states.

The data gathered from the national survey suggest that statistics graduate students often experience CoPs through required meetings and voluntary discussions about teaching. Participants feel comfortable disagreeing with the people they perceive to be most influential on their teaching beliefs. Most participants perceive a faculty member to have the most influential role in shaping their teaching beliefs.

The survey data did not provide evidence to support the proposed theoretical model relating characteristics of experiences in CoPs and beliefs about teaching statistics. Based on cross-validation results, prior beliefs about teaching statistics was the best predictor of current beliefs. Additional models were retained that included student characteristics suggested by previous literature to be associated with student-centered or traditional teaching beliefs (e.g., prior teaching experience, international student status).

The results of this study can be used to inform future efforts to help promote student-centered teaching beliefs and teaching practices among statistics GTAs. Modifications to the GETS Inventory are suggested for use in future research designed to gather information about GTAs, their teaching beliefs, and their experiences in CoPs. Suggestions are also made for aspects of CoPs that might be studied further in order to learn how CoPs can promote teaching beliefs and practices that support student learning.

Table of Contents

Acknowledgements...i

Dedication ii

Abstract iii

List of Tables x

List of Figures xii

Chapter 1 Introduction 1

1.1 Rationale for the Study 2

1.2 Overview of the Dissertation 4

Chapter 2 Review of the Literature..6

2.1 GTAs in Undergraduate Education...7

2.2 GTA Training and Development Programs 8

2.2.1 Empirical studies of GTA training and development programs 9

2.2.2 Teaching Observations of GTAs 10

2.2.3 Mentoring of GTAs..11

2.3 GTAs in Statistics 13

2.4 Teacher Beliefs 15

2.5 Communities of Practice..17

2.5.1 CoPs in primary and secondary education 18

2.5.2 CoPs in Higher education...20

2.5.3 CoPs for GTAs 20

2.6 A Framework for Studying CoPs in Education 22

2.6.1 Engagement 24

2.6.2 Norms of interactions 24

2.6.3 Orientation to practice 25

2.6.4 Leadership presence 25

2.7 Additional Variables Related To GTA Teaching Beliefs and Professional Development 27

2.7.1 International GTAs 27

2.7.2 Prior teaching experience 28

2.7.3 Interest in teaching 29

2.8 Discussion and Critique of the Literature 30

2.8.1 Summary of research on GTA professional development 30

2.8.2 Recommendations for GTA professional development 32

2.8.3 A theoretical model relating experience in CoPs and beliefs 34

2.8.4 Description of the study 35

2.8.5 Research question 35

Chapter 3 Methods 37

3.1 Overview of the Study 37

3.2 Steps in Developing and Revising the Survey 38

3.2.1 Preliminary focus groups 39

3.2.2 Blueprint 40

3.2.3 Item writing 40

3.2.3.1 Constructs 1 & 2: Current and Prior Beliefs 40

3.2.3.2 Construct 3: Engagement 41
3.2.3.3 Construct 4: Norms 41
3.2.3.4 Construct 5: Orientation 42
3.2.3.5 Construct 6: Faculty Presence 42
3.2.3.6 Other characteristics 42
3.2.4 Instrument first draft 43
3.2.5 Initial Feedback 43
3.2.6 Think-Aloud Sessions 44
3.2.7 Final Pilot Session 45
3.3 Survey Instrument 46
3.4 Target Population 48
3.5 Participant Recruitment 50
3.6 Chapter Summary 51
Chapter 4: Results 53
4.1 Description of the Sample 54
4.1.1 Beliefs about teaching statistics 57
4.1.2 Participation in CoPs 64
4.1.3 Norms of interaction in CoP 68
4.1.4 CoP Orientation toward student-centered teaching 69
4.1.5 Faculty presence 71
4.2 Calculation of Measures 73
4.2.1 Measure of current beliefs 73

4.2.4 Measure of prior beliefs 78

4.2.3 Measure of perceived engagement in the CoP 80

4.2.4 Measure of norms of interaction in the CoP 84

4.2.5 Measure of perceived CoP orientation to ward student-centered teaching 88

4.2.6 Measure of perceived faculty presence in CoP 91

4.3 Path Analysis 94

4.3.1 Preparation for cross-validation and model selection 97

4.4 Cross-Validation and Model Selection 99

4.4.1 Cross-validation and model selection: base model 101

4.4.2 Cross-validation and model selection: models with characteristic variables 103

4.4.3 Final (base) model 106

4.5 Chapter Summary 107

Chapter 5 Discussion 108

5.1 Statistics Graduate Students' Beliefs about Teaching Statistics 109

5.2 Statistics Graduate Student Participation in CoPs 111

5.3 The Relationship between Statistics Graduate Students' Experiences in CoPs and Beliefs about Teaching Statistics 113

5.4 The *Graduate Students' Experiences Teaching Statistics Inventory* 115

5.5 Limitations 119

5.6 Implications for Future Research 121

5.6.1 The role of faculty 121

5.6.2 The role of positive norms of interaction 122

5.6.3 Measuring interest in teaching 122

5.6.4 Measuring experiences of CoPs 123

5.6.5 Measuring teaching beliefs and teaching practices 123

5.7 Conclusion 125

References 126

Appendix A: Theoretical Model Relating Graduate Students' Experiences of CoPs and Their Beliefs about Teaching Statistics 139

Appendix B: GETS Inventory Survey Instrument and Response Summaries 140

Appendix C: Focus Group Questions 154

Appendix D: Survey Blueprint 156

Appendix E: Examples of Original and Revised Items 160

Appendix F: Solicitation for Participation Posted in the Caucus of Academic Representatives Weekly Digest 163

Appendix G: First E-mail Solicitation for Participation Sent to CAR 165

Appendix H: Second E-mail Solicitation for Participation Sent to CAR 167

Appendix I: Example of First E-mail Solicitation to Faculty Contacts 170

Appendix J: Example of Follow-up E-mail Solicitation to Faculty Contacts 172

Appendix K: Institutions Represented in Sample Used to Compute Final Model 174

Appendix L: Correlation Matrix of Scores for the Six Core Constructs 176

Appendix M: R Code for Cross-Validation and Model Selection 177

List of Tables

Table 1 *Four Aspects of CoPs members' Experiences of CoPs* 23

Table 2 *CoP Perspective on Recommendations for Statistics GTA Professional Development* 33

Table 3 *GETS Inventory Items by Section* 47

Table 4 *Description of the Sample* 55

Table 5 *Quantitative Characteristics of the Participants Whose Responses Were Used in the Final Model* 57

Table 6 *Current and Prior Beliefs About Teaching Statistics* 58

Table 7 *Pairwise-Complete Intercorrelations of Topics Regarding Current Beliefs about Teaching Statistics* 62

Table 8 *Pairwise-Complete Intercorrelations for Topics Regarding Prior Beliefs about Teaching Statistics* 63

Table 9 *Counts (and Percentages, Calculated Within Level of Teaching Responsibilities) of Participation in Meetings about Teaching, Grouped by Highest Experience Level of Teaching Responsibility* 65

Table 10 *Five Number Summaries of Variables Collected to Indicate Level of Participation in Communities of Practice* 66

Table 11 *Intercorrelations for Variables Designed to Measure Engagement in Communities of Practice*..67

Table 12 *Norms of Interaction with the Two Department Members Chosen as the Most Influential Upon Participants' Teaching Beliefs* 69

Table 13 *Orientation Toward Student-Centered Teaching of the Department Member Who is Perceived to be Most Influential Upon Teaching Beliefs* 70

Table 14 *Polychoric Intercorrelations of Topics Designed to Measure CoP Orientation Toward Student-Centered Teaching* 71

Table 15 *Summaries of Faculty Attendance of Voluntary and Required Meetings* 72

Table 16 *Confirmatory Factor Analysis Results for Models of the Current Beliefs Measure* 76

Table 17 *Coefficient Estimates and Fit Measures for Models of Prior Beliefs Measure*..79

Table 18 *Item CFA Coefficients and Fit Measures for Models of the Measure of Engagement in the CoP* 83

Table 19 *Coefficient Estimates and Fit Measures for Models of Norms of Interaction Measure* 86

Table 20 *Coefficient Estimates and Fit Measures for Models of CoP Orientation Measure*..88

Table 21 *Coefficient Estimates and Fit Measures for Models of Faculty Participation Measure*..93

Table 22 *Path Analysis Coefficients and Fit Measures* 96

Table 23 *Cross-Validation Results of 1000 Resplittings for Candidate Base Models* .102

Table 24 *Two-fold Cross-Validation Results for j=1000 resplittings of Models With Characteristic Variables* 105

List of Figures

Figure 1. Theoretical model relating four aspects of CoPs to current and prior teaching beliefs 35

Figure 2. Changes in Beliefs Since Entering Current Degree Program 60

Figure 3. Distribution of scores for the measure of Current Beliefs 78

Figure 4. Distribution of scores for the measure of Prior Beliefs 80

Figure 5. Scores for measure of Engagement in CoP 84

Figure 6. Distribution of Scores for Norms of Interaction in CoP 87

Figure 7. Distribution of scores for measure of perceived Orientation of CoP 90

Figure 8. Orientation scores conditioned upon role of most influential person (faculty or not faculty) 90

Figure 9. Distribution of scores for measure of perceived Faculty Presence in CoP 94

Figure 10. Plot of (jittered) Current Beliefs vs. (jittered) Prior Beliefs, with Model 1 overlaid 106

Figure 11. Theoretical model relating four aspects of CoPs to current and prior teaching beliefs with path coefficients 139

Chapter 1

Introduction

Worldwide, graduate students can be found teaching and assisting with courses at post-secondary institutions. It is advantageous for departments to employ graduate students in these roles because they cost less than full-time faculty (Muzaka, 2005). Also, graduate students are convenient part-time employees whose availability typically aligns with the institution's academic terms. From the graduate students' perspective, teaching experiences can be included on resumes and improve their chances for future employability. This opportunity is particularly attractive for graduate students who wish to pursue careers in academia. In addition, many graduate students receive tuition benefits or a stipend in exchange for their services related to teaching.

Institutions hire graduate students in many different positions related to teaching. Svinicki (1989) suggests that the teaching tasks that graduate students perform are typically some of the most demanding in the teaching profession. In Canada and the United States, graduate students may have the entire responsibility for designing and administering courses. Other graduate students are hired for assisting courses by grading assigned work, holding office hours, and monitoring course websites (Hoessler & Godden, 2015; Park & Ramos, 2002). In the United Kingdom, graduate students in science departments have been found to facilitate labs and fieldwork, whereas those in social sciences and humanities lead discussion groups.

For the purposes of this paper, graduate teaching assistants (GTAs) will be defined as graduate students who have been hired to be the instructor of record, or to assist the instructor of record in any way associated with teaching undergraduate or graduate courses. This definition includes many different titles used across the world including *teaching fellows*, *teaching assistants*, *moniteurs*, and department-hired *graduate tutors*.

1.1 Rationale for the Study

Statistics departments in the United States appear to rely heavily upon GTAs to teach introductory statistics courses. Blair, Kirkman, and Maxwell (2013) found that GTAs teach 25% of introductory courses in departments that grant doctoral degrees in statistics. Departments are likely to continue to rely upon GTAs to teach statistics as a way to reduce costs while maintaining course enrollment (e.g., Birch & Morgan, 2005).

Unfortunately, many statistics GTAs appear to hold teaching beliefs and teaching practices that are not aligned with current recommendations for teaching statistics. One way to categorize teaching beliefs and practices is on a spectrum from teacher-centered to student-centered beliefs and practices (Kember, 1997). Teacher-centered teaching beliefs and practices focus on transferring structured knowledge to students. In contrast, student-centered teaching beliefs and practices focus on facilitating understanding and fostering conceptual change. Student-centered teaching beliefs have been endorsed by the American Statistical Association (ASA, 2005, 2016) and the American Association for the Advancement of Science (AAAS, 1989). However, a recent study found that many statistics GTAs ascribe to teaching beliefs and teaching practices that are not student-centered (Justice, Zieffler, & Garfield, in press).

A variety of professional development experiences have been created to try to cultivate student-centered teaching beliefs and practices among statistics GTAs (e.g., Garfield & Everson, 2009; Rumsey, 1998). A special section of *The American Statistician* in 2005 was dedicated to this topic, and in this issue several departments shared their strategies for preparing GTAs to teach (e.g., Gelman, 2005). However, little empirical evidence has been collected to explore the extent to which the professional development strategies are effective.

Unfortunately teaching beliefs can be very resistant to change, and can affect—or even impede—teachers' experiences of professional development opportunities (e.g., Borko & Putnam, 1995; Pajares, 1992). Teaching beliefs can also be very difficult to measure and difficult to study (e.g., Fang, 1996). Some researchers believe that studies of teacher beliefs should take into account the cultures and contexts surrounding the professional development (e.g., Putnam & Borko, 2000). This approach may be particularly important for research designed to study GTA professional development related to teaching. Evidence suggests that GTAs appear to be particularly influenced by each other. GTAs have been found to rate interactions with each other as the most valuable and seek information from each other first (e.g., Darling, 1987; Myers, 1994).

One approach for examining GTA professional development is communities of practice. Communities of practice may be defined as groups of practitioners who deepen their knowledge and expertise by interacting on an ongoing basis (Lave & Wenger, 1991; Wenger, McDermott, & Snyder, 2002). Communities of practice offer a natural perspective for studying the effects of professional development opportunities upon GTA beliefs because they incorporate the cultures and contexts surrounding professional development activities. GTA's interactions with one another are a form of participation in the community of practice. This study seeks to explore the relationship between statistics graduate students' experiences in communities of practice and the statistics teaching beliefs that they hold.

1.2 Overview of the Dissertation

Chapter 2 offers a review of literature related to this study. To establish foundations for GTA professional development, the chapter includes summaries of studies related to GTAs in statistics (specifically), as well as studies related to GTAs in all disciplines. A brief overview of literature related to teacher beliefs is also given. Based on the literature, communities of practice are introduced as one way for studying the professional development of GTAs. Background information regarding communities of practice is offered, with particular focus on communities of practice in education settings. Connections are made between studies and a four-part framework for studying GTA communities of practice is offered. The framework is used to create a theoretical model for the relationship between GTAs' beliefs and their experiences in communities of practice. The research question for the study involves testing and refining the model.

Chapter 3 describes the methodology for the study. The survey development process is described and an overview of the final survey instrument is included. The chapter also includes a description of the three methods used for recruiting graduate students to participate in the survey. The target population for the study and the actual population from which participants were drawn are defined.

Chapter 4 gives the results of the data collection and analyses. A description of the sample is given, including demographic characteristics of interest. Results regarding participants' beliefs about teaching statistics and their participation in communities of practice are presented. Based on the results, measures of the six core constructs of the study are defined. The measures are used to explore the theoretical model offered in Chapter 3, as well as other theoretical models relating participants' beliefs about teaching statistics and their participation in communities of practice. Analysis is conducted using cross-validation methods to protect against overfitting. Finally, models that include characteristic variables are explored to examine model invariance across different student populations.

Chapter 5 offers a discussion of the results. The discussion highlights important results regarding participants' beliefs about teaching statistics and participants' perceived experiences in communities of practice. The relationship between participants' perceived experiences in communities of practice and their beliefs about teaching statistics is discussed as well. The chapter also includes an overview of items that were used in the survey and which may be useful in future studies for measuring teaching beliefs and participation in communities of practice. The chapter closes with limitations and implications for future research.

Chapter 2

Review of the Literature

The purpose of this study is to investigate how variation in graduate students' perceptions of their experiences in communities of practice may be related to their teaching beliefs. Also of interest is exploration of whether the relationships are similar for different groups of graduate students (e.g., international GTAs vs. native GTAs; GTAs with prior teaching experience).

To provide background for the study, this chapter offers a review of relevant literature. Much of the literature regarding graduate students' beliefs about teaching focuses on GTAs. After a brief review of studies related to GTAs in undergraduate education, the next section reviews literature about strategies that have been used to prepare GTAs for teaching. A summary is offered of literature related to GTAs in statistics, specifically. A brief review of literature related to teacher beliefs is also given. The section that follows reviews an alternative approach to studying GTA professional development related to teaching—namely, a communities of practice approach.

A discussion and critique of the literature is offered next. Based on literature regarding the study of communities of practice, a four-construct framework for studying communities of practice is given. A theoretical model is posited indicating relationships between graduate students' beliefs and the four constructs of the communities of practice framework. Finally, the research question for the study is given, namely, to explore these relationships.

2.1 GTAs in Undergraduate Education

Multiple studies have found that GTAs are responsible for more than one-fifth of undergraduate course credits. Buerkel-Rothfuss and Gray (1990) suggest that departments across all disciplines, on average, rely on GTAs to generate about 20 percent of university credit hours. GTAs were the sole instructor for two-thirds of these credit hours. Nyquist and Wulff (1987) identified similar patterns in at the University of Washington and eight peer institutions, where GTAs were responsible for about 25 percent (or more) of all undergraduate instruction in 1980. GTAs appear to be particularly responsible for introductory course credits (Nyquist & Wulff, 1987).

Survey studies indicate that GTAs have concerns related to their preparation for teaching. Based on a survey of over 4,000 PhD students from 27 institutions, Golde and Dore (2001) found that GTAs did not feel prepared for their teaching roles (e.g., less than half of the respondents felt equipped to grade assignments fairly). Results of a survey of more than 30,000 doctoral students suggest that 45 percent of surveyed GTAs felt they had received insufficient training for their service as instructors or teaching assistants (Fagen & Suedkamp Wells, 2004). Nearly half of respondents lacked faculty supervision for improving their teaching. In another national study, Diamond and Gray (1987) found that GTAs would like more preparation for teaching tasks (e.g., conducting classroom discussions, preparing tests, evaluating one's teaching). GTAs also indicated they would like more faculty support for teaching.

2.2 GTA Training and Development Programs.

The literature has much to say about programs designed to train and develop GTAs for teaching. There are edited volumes (e.g., Marincovich, Prostko, & Stout, 1998; Nyquist, Abbott, Wulff, & Sprague, 1991; Chism, 1989; Wulff & Austin, 2004), proceedings from national conferences (e.g., 1986 Conference on Institutional Responsibilities and Responses in the Employment and Education of Teaching Assistants; Columbus, Ohio), and even a journal dedicated to the topic (*Journal of Graduate Teaching Assistant Development,* established in 1993).

Different types of training and development programs have been created to prepare GTAs for their teaching responsibilities. Some programs are university-wide and intended for GTAs in all disciplines (e.g., Wulff, Nyquist, & Abbott, 1991; Schoem, Carlton, Gates & Black, 1991). Other programs are discipline-specific (e.g., Fernald, 1995; Hammrich, 1996; Speer, 2004; Wyse, 2010), or combinations of both university wide and discipline-specific components (e.g., Jones, 1993; Luft, Kurdziel, Roehrig, & Turner, 2004). Programs vary in length, in GTA participation expectations (optional or required), GTA compensation, curricular emphasis, and practice opportunities (Parrett, 1987). Programs also vary by evaluation techniques and the extent to which follow-up activities are used (Weimer et al., 1989).

Some of the literature on GTA development programs is descriptive, sharing ideas through example training programs (e.g., Cahyadi & Butler, 2005; Davis & Minnis, 1993, Hammrich 1996; Nyquist & Wulff, 1987; Schoem et al., 1991). In a review of literature in GTA training, Carroll (1980) called for more empirical research on the effects of training programs, rather than sharing of innovative ideas. Years later, reviews by Parrett

(1987) and Abbott, Wulff, and Szego (1989) still echoed Carroll's call for data-supported research on effects of programs on GTAs. To date, there are still relatively few empirical studies of the effects of GTA training and development programs.

2.2.1 Empirical studies of GTA training and development programs. The empirical research on GTA development related to teaching is often subject to methodological concerns. For some studies control groups are not included (e.g., Boman, 2013). The results of these studies are subject to confounding variables (e.g., growth over time, increased experience teaching in the classrooms while the study was being conducted). Some studies may have been too rushed to discern effects. In a yearlong study of Biology GTAs, Wyse (2010) found that some effects take more than one semester to be detected. In other studies issues arise regarding the psychometric properties of instruments used to measure outcomes. All in all, it is difficult to trust many of the significant and insignificant results that have been reported regarding GTA professional development strategies.

Empirical studies of GTA development programs have explored four main response variables: student variables (e.g., student achievement, student ratings of GTAs); GTA teaching affect; GTA teaching practices; and GTA teaching beliefs (e.g., Abbot, Wulff, & Szego, 1989; Boman, 2013; Bray & Howard, 1980; Carroll, 1980; Dalgaard, 1982; Gilmore et al., 2013; Prieto & Altmaier, 1994; Rodriques & Bond-Robinson, 2006; Roehrig, Luft, Kurdziel & Turner, 2003; Saroyan, Dagenais, & Zhou, 2009; Shannon, Twale, & Moore, 1998; Williams, 1991; Wyse, 2010). For the former two outcomes (student variables and GTA teaching affect), studies and reviews have provided fair evidence of associations and perhaps even causal improvements after training. The effects of training programs on GTAs' teaching practices and teaching beliefs are less clear. While some training programs have been associated with improved teaching practices and more student-centered teaching beliefs, others have not.

The bundling of many program components together makes it difficult to discern whether any effects that are reported are due to one program component, several program components, or interactions between them. There have been calls for investigations of the impact of separate program components (e.g., Abbot, Wulff, & Szego, 1989; Bray & Howard, 1980). Two components that appear particularly promising are teaching observations and mentoring.

2.2.2 Teaching observations of GTAs. GTAs have been observed performing teaching tasks in-person or using video. One form of observation employed in GTA training and development programs is observation of *microteaching*, which is a short practice teaching simulation performed before peers. Typically, microteaching is situated outside of regular classroom teaching. Another form of observation occurs during GTAs' regular teaching sessions.

There is some evidence to suggest that observation of teaching with feedback is a key component for improving teaching behaviors and teaching affect. In a controlled randomized study conducted to investigate the usefulness of various components of a development program, Bray and Howard (1980) assigned one experimental group of GTAs consultations after video observations. They found that teaching behaviors and attitudes toward teaching were significantly improved for GTAs who received consultations over video of their teaching ($p < .001$). It is interesting that significant differences were not found between the GTAs who received video consultations as compared to those who had video consultation in addition to other training components such as seminars and evaluation consultations (p = .589). The authors conclude that the video consultation was the key component of the training program.

Other evidence has been provided to suggest that observations with consultation are a key component of training programs. Williams (1991) found that a training program for English GTAs reduced anxiety only when coupled with observations with consultations and peer GTA mentoring. GTAs experiencing the training program without these components did not show significant reduction in anxiety.

GTAs also appear to value observations with consultation more than other training program components. Dalgaard (1982), found that video consultations with feedback were rated by participant GTAs as the most effective component of a training program that also included topics such as writing assessments, planning lessons, and methods for student-centered teaching.

2.2.3 Mentoring of GTAs. Mentoring is used in many different professions (e.g., medicine, business management). Numerous definitions have been offered (e.g., Anderson & Shannon, 1988; McKimm, Jollie, & Hatter, 2007). For the purpose of educational research, Healy and Welchert (1990) suggest an operational definition: "a dynamic, reciprocal relationship in a work environment between an advanced career incumbent (mentor) and a beginner (protégé) aimed at promoting the career development of both." (Healy & Welchert, 1990, p. 17).

Experts in elementary, secondary, and higher education suggest many reasons why it is difficult to conduct research on mentoring programs. Little (1990) notes that attention to outputs (e.g., hours spent meeting) rather than outcomes (e.g., depth and meaningfulness of relationships) makes it difficult to compare or assess reported results of mentoring programs. Little also argues that mentors are often selected based on classroom performance, not mentoring skills (which are very different from each other). After a multi-year, government-funded study on mentors in higher education, Boice (1992) found that mentors viewed protégés as colleagues and did not want to impose authority over them. He also found that protégés with mentors from their own departments were reluctant to share vulnerably for fear of what evaluations and judgments the mentor might make. In addition, mentor and protégé pairings often found it difficult to continue to find time to meet with one another.

Studies suggest that mentoring may be associated with improved GTA teaching affect, teaching beliefs, and teaching practices. In a controlled study of 27 GTAs in English departments, Williams (1991) found that peer GTA mentoring was one of two components necessary for a training program to show significant reduction in anxiety. Also, in a survey of over sixty GTAs from STEM disciplines, Gilmore et al., (2013) studied correlations between GTAs' beliefs about teaching and four related factors: mentoring, teaching experience, research experience, and training program experience. They found that mentoring was far more influential than all other factors on GTAs' student-centered teaching beliefs. A case study by Volkmann and Zgagacz (2004) recorded changes in a physics GTA's teaching beliefs and practices while engaging in a mentoring relationship. Mentoring has also been employed by the Preparing Future Faculty (PFF) program, which has been cited as "one of the most systematic efforts to increase graduate student preparation for teaching" (Pruitt-Logan & Gaff, 2004).

Many experts have offered critiques of mentoring as a form of teacher professional development. Putnam and Borko (2000) critique mentoring because it can perpetuate prevailing traditional cultures in education. Korpan (2014) argues that faculty mentoring of GTAs does not serve GTAs' immediate needs, is plagued by a large differential in power, places a heavy burden on mentors, and tends to be unidirectional. She echoes the arguments of Lave (1996), who encourages educational researchers to look beyond traditional notions of a single "teacher" and "learner" to find where and how learning occurs. Instead, Korpan recommends larger community of mentors who can help GTAs experience more authentic training. Korpan and Lave's arguments lead toward a communities of practice perspective for studying GTA professional development.

Before reviewing literature related to communities of practice, background will be provided for two other important topics related to this study. The two topics are: GTAs in the discipline of statistics, specifically, and teacher beliefs.

2.3 GTAs in Statistics

To date, there are three empirical studies related to GTAs teaching statistics. All three suggest a need for improved preparation and support for GTAs teaching statistics. First, based on a survey of 68 GTAs from 18 institutions, Noll (2011) found that many GTAs lack essential statistical content knowledge for designing quality instruction or making judgments about the reasonableness of students' answers. Second, in a survey of 213 GTAs representing 38 Ph.D.-granting institutions in the United States, Justice et al. (in press) found that that the majority of surveyed GTAs have not learned about the current recommendations for teaching introductory statistics (e.g., ASA, 2005), nor do they hold teaching beliefs and practices aligned with such recommendations. Third, Green (2010) found that GTAs in one Ph.D.-granting statistics department desired more direction with regards to the use of technology as well as which topics should be emphasized in their courses. These empirical studies suggest that many GTAs in statistics departments need more knowledge, preparation, and support as they fulfill their teaching roles.

The literature related to the development of GTAs in statistics departments is primarily descriptive, sharing ideas and methods that have been used to prepare GTAs for teaching. For example, Garfield and Everson (2009) describe a graduate course for future teachers of statistics that builds on research on the teaching and learning of statistics. Rumsey (1998) describes a collaborative approach to GTA development that uses weekly meetings and seeks to establish a supportive environment. In a special section of the *American Statistician*, strategies for GTA development from four more statistics departments were highlighted. In addition to courses and weekly meetings (e.g., Birch & Morgan, 2005; Gelman, 2005; Harkness & Rosenberger, 2005), strategies included mentoring (Froelich, Duckworth, Stephenson, 2005) and immersion in a departmental culture (Birch & Morgan, 2005).

All of the featured statistics departments' strategies appear to encourage what Kember (1997) describes as teaching that is student-centered (focused upon facilitating conceptual change in students) as opposed to teacher-centered (focused upon the transferring of structured knowledge to students). For example, at Columbia University, GTAs are encouraged to spend less time lecturing (Gelman, 2005). At Penn State, interactive learning is an emphasis of the GTA development program (Harkness &

Rosenberger, 2005). Froelich et al. (2005) point out that in their department "lectures" include small group activities, demonstrations and opportunities for students to actively participate in the class.

Although the strategies described in the descriptive literature are designed to facilitate student-centered teaching, little empirical evidence has been collected to study whether they are successful in achieving change in GTAs' teaching practices or teaching beliefs. Unfortunately, research on the professional development of primary, secondary, and tertiary teachers suggests that changes in teaching practices and teacher beliefs do not come easily.

2.4 Teacher Beliefs

Studies suggest that teacher development experiences can be affected and even impeded by long-held beliefs that are resistant to change (Borko & Putnam, 1995; Calderhead, 1996; Fang, 1996; Kane, Sandretto, & Heath, 2002; Kember, 1997, Pajares, 1992; Simmons et al., 1999; Widen et al., 1998). Beliefs have been found to serve as a filter through which all professional development experiences are perceived. Some researchers go so far as to posit that beliefs must change for teaching practices to change. This research has led to a renewed focus on teacher beliefs. For example, in the *Second Handbook of Research on Mathematics Teaching and Learning*, Sowder (2007) argues that goals of mathematics teacher professional development should involve changing teachers' understandings of how students learn and challenging beliefs that are long-held by teachers.

In a review of studies of tertiary educators' beliefs about teaching mentioned in the introduction to this paper, Kember (1997) found that beliefs can be categorized along a spectrum: teacher-centered versus student-centered. Teacher-centered beliefs tend to focus upon the transferring of structured knowledge to students. On the other end of the spectrum, student-centered beliefs take a more developmental approach; the instructor is viewed as facilitating understanding for the purpose of conceptual change in students. This spectrum has been applied to various aspects of teaching topics. For example, in a recent study of GTA teaching beliefs, Douglas et al. (2016) found that many GTAs' beliefs about content were teacher-centered while their beliefs about student learning were student-centered.

Researchers who wish to study teacher beliefs face many challenges (Calderhead, 1996; Fang, 1996; Kane et al., 2002; Kember, 1997; Pajares, 1992). Teaching beliefs vary in strength and kind, are often grounded in personal experience, and can go unrecognized by the teacher who holds them. The latent nature of beliefs can make them difficult to study, observe, or measure, particularly as teachers must negotiate competing beliefs while making instructional decisions (e.g., Fang, 1996; Lampert, 1985; Pajares, 1992; Razfar, 2012).

Some researchers have proposed that studies of teacher beliefs should take social contexts into account. Cooney (1994) suggests that teachers' beliefs about teaching initially form in social settings and may be reformed only in social settings. Putnam and Borko (2000) draw from social learning theories (e.g., Vygotsky, 1978) to posit that study of teacher development should focus on the effect of the different social settings in which learning is designed to take place. The next section offers a review of literature related to communities of practice, a way of examining teacher professional development that takes social contexts into account.

2.5 Communities of Practice

Communities of practice (henceforth CoPs) are defined as "groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis" (Wenger et al. 2002, p.4). CoPs necessarily involve deepening knowledge and expertise. In this sense CoPs are a form of professional development.

The term *CoP* was coined by Lave and Wenger (1991) as they studied apprenticeship in various contexts. They found that novices' learning was not centralized around a single mentor. Rather, learning occurred through participation in communities of experts and novices (Lave, 1996; Lave & Wenger, 1991; Wenger et al., 2002). Building on Vygotsky (1978), Lave and Wenger introduced the CoP social learning theory that has since been utilized in many fields such as business, industry, and education. According to Lave and Wenger (1991), CoPs are an age-old phenomenon; only the theory and name are new.

Wenger et al. (2002) note the variety of forms that CoPs can take. Although they often have regular structured meetings, CoPs may engage only informally. In fact, CoPs may even exist unrecognized by their host organizations and institutions (Schlager & Fusco, 2003). Wenger et al. offer three characteristics that can be helpful in identifying a CoP. Each CoP has: (1) a specific domain about which the community is focused; (2) trust-filled relationships; and (3) a shared practice that develops over time.

CoPs may serve as an asset to businesses and organizations. For example, participation in CoPs can keep experts interested and at the cutting edge of the practice (Wenger et al., 2002). CoPs can steward valuable knowledge and, under certain conditions, can pass the knowledge on to newcomers. Lave and Wenger (1991) suggest that novices learn only when they are able to participate in the practice. In particular, novice participation should progressively move toward more central tasks to the practice. (They call this type of participation *legitimate peripheral participation*). Lave and Wenger also maintain that aspects of the practice must be transparent to the novices for learning to occur.

Unfortunately, CoPs are not always benevolent to their host organizations (e.g., Lave & Wenger, 1991; Schlager & Fusco, 2003; Wenger et al., 2002). CoPs can be closed to new ideas or promote mediocrity. They may develop over domains that may be damaging or stifle innovation. Core community members may develop such strong trusting relationships that they may not be open to newcomers.

Wenger et al. (2002) claim that it is important for host organizations and institutions to cultivate their CoPs. Cultivation can be a delicate process. CoPs are often resistant to outside forces and can be suffocated. Some methods that have been suggested for cultivating CoPs are: developing and supporting community leadership; providing time and space for the community to meet; offering resources (e.g., refreshments, funding for guest speakers); giving the community voice in the organization; and offering guidance toward the cutting edge of the practice.

2.5.1 CoPs in primary and secondary education. In primary and secondary educational settings, there is a growing body of literature regarding the use of CoPs for teacher professional development. CoPs in education are not always executed according to the vision of Lave and Wenger (1991). For example, CoPs in education are often forced, or a part of a teacher's contract. This feature is not according to the vision of Wenger et al. (2002), who defined CoPs to be voluntary.

There is evidence to suggest that CoPs can lead to positive outcomes in K–12 teaching. In their review of ten empirical studies in K–12 teacher development, Vescio, Ross, and Adams (2008) found positive outcomes associated with CoPs that focus on increased student learning. The outcomes include student-centered teaching practices, higher student proficiency on standardized tests, and improved collaborative cultures in the schools that host the CoPs. A study of primary and secondary professional development in Chicago schools conducted by Smylie, Allensworth, Greenberg, Harris, and Luppescu (2001) found that CoPs may provide motivation, direction, accountability, and feedback to teachers. In addition, Schlager and Fusco (2003) cite evidence to suggest that successful attempts to change teachers toward more reformed curricula are associated with access to CoPs (e.g., Blumenfeld, Fishman, Krajcik, & Marx, 2000; McLaughlin & Mitra, 2001).

Researchers have faced difficulties when attempting to study CoPs in education. In some cases difficulties arise because of the many aliases used for CoPs (e.g., *professional learning communities*, Borko, 2004; *discourse communities*, Putnam & Borko, 2004). Vescio et al. (2008) state that such terms are often overused and misapplied. After a qualitative study of a CoP, Little (2002) identifies several pitfalls for studying CoPs. She recommends that studies of CoPs focus on the norms of the practice, the orientation to practice, and the representation of the practice.

2.5.2 CoPs in Higher education. In higher education, CoPs designed for faculty professional development have been referred to as *faculty learning communities*. Miami University is a leader in hosting faculty learning communities. Starting in 1978, Miami University has documented 81 faculty learning communities that have included over one-third of the entire institution's faculty. Cox (2004) reports some positive outcomes associated with faculty involvement in FLCs, including increased awareness of different teaching and learning styles, significantly improved tenure rates, and greater participation in service to the community and to the university. The success of faculty learning communities at Miami University led to grants supporting the dissemination of the Miami models to other institutions (e.g., Ohio Teaching Advancement Project; Fund for the Improvement of Post Secondary Education Project). In 2004, Cox reported that these projects led to about one hundred new faculty learning communities in over thirty institutions.

Cox describes faculty learning communities as "cross-disciplinary faculty and staff group(s)… who engage in an active, collaborative, year-long program with a curriculum about enhancing teaching and learning and with frequent seminars and activities that provide learning, development, the scholarship of teaching, and community building" (Cox, 2004, p. 8). Cox and colleagues define two types of faculty learning communities: topic-based and cohort-based faculty learning communities (Cox, 2004; Richlin & Cox, 2004).

2.5.3 CoPs for GTAs. For this study, a definition of CoPs for GTAs will be used that builds on Wenger et al. (2002), Cox (2004), and Lave and Wenger (1991). *GTA CoPs* may be viewed as groups of GTAs and faculty members who deepen their knowledge and expertise of teaching by interacting together on an ongoing basis. For the rest of this paper, it will be assumed that the domain of a GTA CoP is teaching, unless otherwise specified. Also, note that in this definition, participation may be voluntary or involuntary.

Two examples of GTA CoPs in have been offered in the literature. Holmes, Ives, and Warren (2013) describe a CoP for GTAs in physics. Their program engages senior GTAs in cooperatively creating and sustaining professional development for newcomer GTAs. Senior GTAs design and run workshops, facilitate course-specific weekly meetings, and mentor newcomers. Another GTA professional development opportunity rooted in CoP theory is offered for engineering GTAs (Crede, BorrEgo, & McNair, 2010). In this CoP, GTAs engage in weekly meetings, mentor one another, provide feedback to one another, and participate in gradually increasing practices within the community. After conducting interviews with participants in a qualitative study of this CoP, Crede et al. (2010) found that GTAs desired even more interaction with one another.

There is also evidence to suggest that GTAs in statistics departments participate in CoPs. For example, Rumsey (1998) uses a cooperative teaching approach to develop teaching and leadership among GTAs in her statistics department. In her model, weekly meetings serve as a forum for GTAs to test new ideas and discuss topics related to statistics and pedagogy. Gelman (2005) also includes weekly discussion of teaching topics as part of a course to develop GTAs as statistics teachers at the college level.

With such evidence to suggest that participation in CoPs may be able to serve as professional development for teachers at elementary, secondary, and college levels, the question arises as to the extent to which GTA participation in CoPs may be associated with positive results. To study this question, the following section examines literature related to frameworks for defining and measuring CoPs.

2.6 A Framework for Studying CoPs in Education

Literature related to CoPs in many education contexts (e.g., secondary education, online learning) identifies important characteristics of members' experiences of CoPs. For example, based on a series of qualitative studies of CoPs for teachers in secondary education, Little (2002) offers a three-part framework for studying CoPs. The components of her framework are called the representation of the practice, the orientation to practice, and the norms of interaction. Garrison, Anderson, and Archer (2000; 2010) offer another three-part framework for studying CoPs. Their framework is based on extensive research on students' experiences of online learning. The components of their framework are called the cognitive presence, social presence, and teaching presence.

Connections can be drawn across studies' frameworks. For example, the representation characteristic proposed by Little (2002) can be matched with Lave and Wenger's (1991) concept of masters' transparency, as well as Garrison's et al. (2010) concept of teacher presence. By matching components from the studies, four main aspects of CoPs arise. Table 1 summarizes the aspects of CoPs and their sources. The four aspects are described in greater detail in the following subsections.

Table 1.

Four Aspects of CoPs members' Experiences of CoPs

Aspect	Brief Description	Support in the Literature		
		Author (Date)	Context	Alias(es) used
Engagement	Extent to which practice is cognitively engaging or to which participants are actively involved	Lave & Wenger (1991)	Apprenticeship	Peripheral Participation
		Garrison et al. (2000)	Communities of Inquiry	Cognitive Presence
		Rovai (2002)	Online Learning	Learning Factor
Norms of Interactions	Norms of personal interactions among CoP members (e.g., conversation conventions, openness of communication, participation structures)	Lave & Wenger (1991)	Apprenticeship	Legitimate Participation
		Garrison et al. (2000), Arbaugh et al. (2008)	Online Learning & Communities of Inquiry	Social Presence
		Wenger et al. (2002)	Business & Organization	Trust-filled relationships
		Little (2002)	K–12 Teacher Development	Norms of Interaction
		Rovai (2002)	Online Learning	Connectedness Factor
Orientation to Practice	Domain about which the CoP centers (i.e., central topic of the CoP)	Lave & Wenger (1991)	Apprenticeship	Shared Practice
		Wenger et al. (2002)	Business & Organization	Domain of the CoP
		Little (2002)	K–12 Teacher Development	Orientation to Practice
Leadership Presence	Extent to which leadership is perceived to be present; transparency of masters' practice	Lave & Wenger (1991)	Apprenticeship	Masters' Transparency
		Garrison et al. (2000)	Online Learning & Communities of Inquiry	Teaching Presence
		Wenger et al. (2002)	Business & Organization	Leaders, CoP cultivators
		Little (2002)	K–12 Teacher Development	Representation, Transparency

Note. CoP refers to community of practice (Lave & Wenger, 1991).

2.6.1 Engagement. The first aspect of members' experiences of CoPs that arises in the literature is related to the duration and quality of participants' engagement in a CoP. This aspect of CoPs arises from a many studies of CoPs in other education contexts. For example, Garrison et al. (2000) identified cognitive engagement, including exploration and application of concepts presented by the CoP, as an important part of their framework. In another study of communities in online learning, Rovai (2002) found empirical evidence to support a factor called *learning*, which was loaded upon by items related to desire to learn and satisfaction related to opportunities for learning. The original researchers who coined the term *Community of Practice* insist that learning occurs when members have legitimate participation in the community (Lave & Wenger, 1991), which can be interpreted as a form of engagement.

2.6.2 Norms of interactions. The second aspect of members' experiences of CoPs that will be used for this study is called the norms of interaction. This aspect involves the extent to which CoP participants' interactions with one another are healthy, respectful, and encourage trust. The name of this aspect comes from Little (2002), who suggested that studies of CoPs concern themselves with the nature of face-to-face interchanges and "how conversational conventions, participation structures, and the enacted norms of professional practice open up or close off possibilities for practice and for inquiry into practice," (Little, 2002, p. 936).

The norms of interaction aspect of CoPs arises from other studies, as well. For example, based on social learning theories and factor analysis, Garrison et al. (2000) included *Social Presence* in their framework for describing CoPs. Arbaugh et al. (2008) proposed a framework that involved the extent to which CoP members trust one another and communicate purposefully. Wenger et al. (2010) also suggested that a vital component of CoPs is a set of trust-filled relationships.

2.6.3 Orientation to practice. Another aspect of experiences of CoPs offered by Little (2002) is the CoP's orientation to practice. In Little's definition, this characteristic of CoPs is related to the extent to which the CoP is open to change or interested in improving their shared practice. This aspect of members' experience of CoPs is akin to what Wenger et al. (2002) calls the "domain" of the CoP, which involves the topics that the CoP tends to discuss, the stances the CoP tends to take on issues of the practice, and the aspects of the practice that the CoP endorses. Wenger warns that the domain should not be assumed to be benevolent. Domains of CoPs may be positive or they may reinforce apathy, the status quo, social injustices, or poor practices.

For this study, the aspect of practice that is of most interest is the orientation toward student-centered versus teacher-centered teaching (Kember, 1997). For the purposes of this paper, the *CoP Orientation* will refer to the extent to which the CoP is in support of student-centered teaching.

2.6.4 Leadership presence. The fourth and final aspect of CoPs that arises from connections in literature is the extent to which CoP members perceive leadership to be present. Leadership presence has empirical support from two separate studies of online learning. Garrison et al. (2010) and Arbaugh et al. (2008) both found that *Teacher Presence* was a factor in students' experiences of online learning. In addition to this empirical support, other authors have included constructs related to leadership presence in their observations of CoPs. Lave and Wenger (1991) discuss a vital aspect of healthy CoPs is the transparency of the masters' practice to the novices. Without the masters engaged in the community, novice learning suffered. Wenger et al. (2002) further suggest that CoPs can be cultivated by strategic leadership. Although Little (2002) did not explicitly mention leadership in her three-part framework for studying CoPs, she points out the important role of the *Representation* of the practice. The representation includes how the practice is shared and the transparency of the practice, two aspects that often involve CoP leadership.

Although graduate students can take leadership roles, for the purposes of this study the leadership presence will focus on faculty leadership. There are three reasons for this decision. First, in the context of graduate teaching assistant development, literature suggests faculty can be valuable and perhaps influential leaders. Faculty mentors have been rated as helpful by GTAs (e.g., Jones, 1993). Also, faculty teaching observations have been found to be associated with more student-centered teaching methods (Dalgaard, 1982). Secondly, there is evidence that faculty can influence the topics discussed in CoPs. When Dotger (2011) studied science GTAs involved in a Japanese Lesson Study, she found that discussion had higher quality when faculty members were present. Thirdly, criticism of faculty presence (or lack thereof) in graduate students' training for teaching (e.g., Fagen & Suedcamp Wells, 2004) often appears similar to criticism masters' transparency (or lack thereof) in CoPs (e.g., Lave & Wenger, 1991). For the purposes of this study, the Leadership Presence aspect of members' experiences of CoPs will be referred to as *Faculty Presence*.

2.7 Additional Variables Related to GTA Teaching Beliefs and Professional Development

Studies have been conducted to explore what characteristics of GTAs may be associated with various teaching beliefs, teaching practices, and student outcomes. The three characteristics that will be reviewed in this section are GTAs' international student status, GTAs' prior teaching experience, and GTAs' interest in teaching.

2.7.1 International GTAs. A survey by Davis (1991) found that 12% of U.S. graduate students are foreign. For many foreign graduate students, teaching assistantships are the only legally available employment (Ford, Gappa, Wendorff & Wright, 1991). The literature reports mixed responses to international students' service as GTAs in the United States. By the 1980's concern about international student instructors were so prevalent that the phrase had emerged in the literature: *the foreign TA problem* (e.g., Fisher, 1985; Young, 1989). By 1989, twenty U.S. states had legislature mandating the testing and screening of international GTAs for English proficiency (Thomas & Monoson, 1993).

In response to the concerns about international GTA instructors, much has been written related to the development of international students as GTAs (e.g., Byrd, Constantinides, & Pennington, 1989; Davis, 1991; Ford et al., 1991). A special section of the journal *English for Specific Purposes* (1989) was devoted to the topic. The studies on international GTAs have found evidence to suggest that English Proficiency may not be as important as originally suspected. Some studies have found that factors more important than proficiency include accent (Bailey, 1983; Jacobs & Friedman, 1988;

Rubin, 1992), and culturally sensitive teaching styles (Hoekje & Williams, 1992; Luo, Grady, & Bellows, 2001).

Different types of professional development strategies for international GTAs have been offered (e.g., Constantinides 1987, 1989; Gilreath & Slater, 1994; Sarkisian & Maurer, 1998; Travers, 1989; Weimer, Svinicki, & Bauer, 1989). Some aspects of training programs are agreed upon; others are not. Experts seem to agree that development programs ought to include cultural issues, such as the departure from authoritarian teaching styles. There is less consensus about whether international GTA training programs should be integrated with programs for native GTAs. International GTAs' occasional reluctance to participate in general GTA training programs poses questions about the extent to which international GTAs engage in development programs designed primarily for native GTAs. This raises questions international students' participation in CoPs.

2.7.2 Prior teaching experience. GTAs' prior teaching experience is another characteristic that may be associated with various teaching outcomes and experiences in CoPs. Two correlational studies found significant relationships between self-efficacy and prior teaching experience. However both studies are subject to confounding variables such as age and participation in training programs (Prieto & Altmaier, 1994; Prieto & Meyers, 1999). Regarding the effect of prior teaching experience on teaching behaviors, two studies found no effect (Dalgaard, 1982; Gilmore, Maher, Feldon & Timmerman, 2013). In addition, Boman (2013) found that prior teaching experience had no significant effect on a GTA's ability to improve teaching behaviors after training.

Perhaps the most surprising results are regarding the effect of teaching experience on student evaluations of GTAs' teaching. In a study of 129 GTAs representing 29 departments in one institution, Shannon et al. (1998) found significantly worse student evaluation scores were associated with GTAs who had prior experience as GTAs. In fact, GTAs with no experience at all had higher student evaluation scores than those with previous experience as GTAs ($p < .01$). This result provides some evidence to suggest that that the experience of teaching as a GTA is not associated with improved student evaluations of teaching. GTAs reporting prior elementary, secondary, or college teaching experience (that is, experience teaching not as GTAs) received significantly higher ratings than those with no experience.

2.7.3 Interest in teaching. Another variable worth considering is GTAs' interest in teaching. It is plausible that GTAs who are more interested in teaching may be more engaged and receptive to ideas presented in professional development experiences. There is evidence to suggest that GTAs' interest in teaching may be challenged by general feelings that research is valued more than teaching, and that research experiences contribute more to professional growth than teaching experiences (e.g., Boehrer & Sarkisian, 1985; Ethington & Pisani, 1993; Luft et al., 2004). Hartnett and Katz (1977) suggested a potential source of these attitudes, namely that GTAs are trained by professors to value research above all other activities, including teaching.

It is difficult to find research related to the effects of interest in teaching on teaching beliefs and teaching practices. Of studies that have been conducted to explore attitudes toward teaching statistics, the primary emphasis has been on pre-service teachers' attitudes toward teaching statistics at primary levels (e.g., Hannigan, Gill, &

Leavey, 2013; Martins, Nascimento, & Estrada, 2012). The emphasis at the tertiary level has been on students, not faculty or instructors. The studies tend to measure teachers' interest in statistics, not statisticians' interest in teaching.

Instruments have been designed to measure attitudes toward statistics (e.g., *Attitudes Toward Statistics,* ATS, Wise, 1985; *Survey of Attitudes Towards Statistics,* SATS, Schau, Stevens, Dauphinee, & Vecchio, 1995). These instruments have occasionally been adapted to be administered to pre-service and in-service teachers (e.g., Martins et al., 2012). Inventories have been developed to measure interest in various careers, including teaching (e.g., Strong Interest Inventory). However, these instruments are often proprietary.

Having outlined variables to consider for a framework for studying CoPs and other GTA characteristics that may be important to consider, this review now offers a final discussion of the literature, critique, and formulation of the research question for the study.

2.8 Discussion and Critique of the Literature

The final section of this review offers a discussion and critique of the literature. The critique leads to the formulation of the research question for the study, which includes testing a theoretical model that relates experiences of CoPs with teaching beliefs (student-centered versus teacher-centered; Kember, 1997).

2.8.1 Summary of research on GTA professional development. The majority of literature regarding GTA preparation and development is descriptive. Calls have been made for more rigorous empirical studies (e.g., Carroll, 1980). Of the little empirical research that has been conducted, the results are not entirely conclusive, often due to methodological concerns. The literature related to GTA professional development for teaching also suffers from frequent "bundling" of interventions together. For these studies it is difficult to assess whether the results are attributable to a single intervention or an interaction of many interventions that were bundled together.

There appear to be two major sources of professional development that can be parsed out to have consistent and strong empirical support: mentoring and teaching observations. Evidence based on controlled randomized designs has been found to suggest that these strategies are able to change GTA beliefs and teaching practices (e.g., Williams, 1991). Also, these two professional development strategies are consistently given high ratings by GTAs in terms of helpfulness.

Studies of GTA and teacher professional development have come under criticism for missing an important part of the picture. Researchers suggest that studies of professional development must involve the social contexts surrounding the professional development (e.g., Lave, 1996; Putnam & Borko, 2000;). Social contexts may be particularly important for GTAs (Austin, 2002; Crede et al., 2010; Darling, 1987; Darling & Staton, 1989; Jones, 1993; Myers, 1994; Myers, 1998; Williams & Roach, 1992; Wulff et al., 2004). GTAs look to one another, primarily, to seek information regarding their teaching responsibilities. GTAs find one-another more helpful than any of faculty, workshops, courses, or other sources of professional development. It has even been claimed that GTAs are the strongest influence on one another. Most of the research on programs to develop GTAs for teaching has not taken social contexts into account.

One way of studying GTA professional development that does take social contexts into account is communities of practice. Participation in communities of practice includes structured professional development opportunities (e.g., courses, workshops), as well as informal interactions that occur between members (e.g., unplanned interchanges in offices, e-mail conversations). Communities of practice approaches to GTA professional development have been used in some disciplines including physics (e.g., Holmes et al., 2010), engineering (e.g., Crede et al., 2013), and statistics (e.g., Rumsey, 1998). Little empirical evidence has been collected to discover how a communities of practice approach to GTA professional development is able to facilitate change in teaching beliefs and teaching practices.

2.8.2 Recommendations for GTA professional development. Based on empirical research on GTA preparation and scholarship about statistics GTA preparation, recommendations and implications can be suggested for GTA professional development programs related to teaching. The recommendations and implications can be categorized in three broad themes: provide a community of support for GTAs; provide appropriate experiences for improving teaching with timely feedback; and provide opportunities for increased knowledge related to teaching. Examples of each recommendation are given in Table 2.

The recommendations from the research can also be viewed from a CoP perspective. Column three of Table 2 indicates how each of the recommendations can be satisfied from a CoP perspective of professional development. The table illustrates how recommendations for GTA professional development can be summarized as cultivating healthy CoPs that engage newcomers.

Table 2.

CoP Perspective on Recommendations for Statistics GTA Professional Development

Recommendation	Examples	CoP perspective
Provide a community of support for GTAs	Include a strong component of faculty support for teaching (Belnap, 2005; Diamond & Gray, 1987; Green, 2010) Provide mentoring opportunities with expert teachers. (Green, 2010; Hogg, 1991; Moore, 2005, Noll, 2011; Williams, 1991) Promote relationships between novice & experienced GTAs (Staton & Darling, 1989; Williams & Roach, 1992). Foster community among all GTAs (Darling, 1987; Green, 2010). Create regular support sessions (e.g. weekly meetings) (Green, 2010).	Novices gain authentic, robust learning by engaging with experts and other novices in CoPs.
Provide appropriate experiences for improving teaching with timely feedback	Conduct trial teaching sessions (either by video or in-person) followed by feedback (Abbott et al., 1989; Belnap, 2005, Bray & Howard, 1980). Assign appropriate responsibilities according to each GTA's individual readiness (Kurdziel et al., 2003; Nyquist & Sprague, 1998). Specify responsibilities (Green, 2010). Monitor GTAs proactively & provide feedback (e.g., observations) (Jones, 1993; Wulff et al., 2004).	Learning occurs while engaging legitimately in the practice (not cognitive acquisition). As novices become experts they engage in more central tasks of the practice.
Provide opportunities for increased knowledge related to teaching	Develop content & pedagogical content knowledge for teaching statistics including the selection and ordering of content topics (Green, 2010; Noll, 2011). Guide GTAs about how to use technology to enhance course content (Golde & Dore, 2000; Green, 2010). Prepare ITAs for cultural & linguistic differences (Luo et al., 2001).	The CoP stewards knowledge for how the practice is conducted, and shares the knowledge with novices through participation in the CoP.

Note. CoP refers to community of practice (Lave & Wenger, 1991).

2.8.3 A theoretical model relating experiences in CoPs and beliefs. Participation in CoPs has been shown to be associated with changes in beliefs in education research at primary, secondary, and college levels (e.g., Blumenfeld et al., 2000; Cox, 2004; McLaughlin & Mitra, 2001; Vescio et al. 2008). To study the extent to which experiences in CoPs may be associated with student-centered beliefs for GTAs, it is important to determine what aspects of CoPs should be studied and how CoPs can be characterized.

Based on a synthesis of literature on CoPs in education and other settings, this review offered four aspects of members' experiences in CoPs. The four aspects are: Faculty Presence, CoP Norms of Interaction, the CoP's Orientation to Practice, and Engagement in the CoP. Some of the aspects may be related to one another (e.g., it is plausible that GTAs are more engaged when the CoP has more faculty presence and healthier norms of interaction). Also, there may be associations between the aspects and the extent to which graduate students' beliefs about teaching statistics are student-centered.

Figure 1 offers a theoretical model relating graduate students' beliefs about teaching statistics and their experiences of the four aspects of CoPs used in this study. In the spirit of Keith (2006), the choices of paths and their directions were guided by results of prior studies, theory, time precedence, logic, and parsimony. Also as suggested by Keith, prior teaching beliefs have been included in the model, because they are a plausible common cause of teaching beliefs, teaching practices, and CoP orientations toward student-centered teaching.

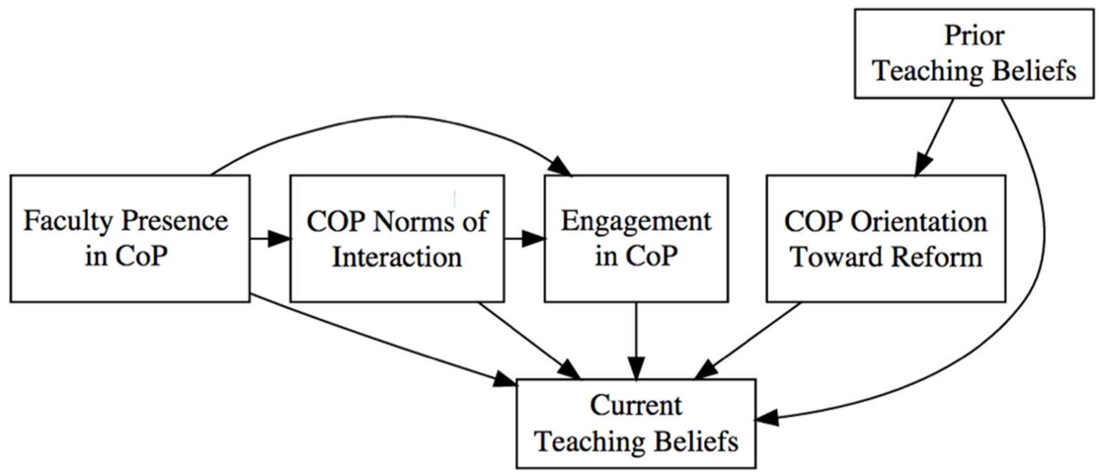


Figure 1. Theoretical model relating four aspects of CoPs to current and prior teaching beliefs.

2.8.4 Description of the study. This study is motivated by research that suggests graduate students need more professional development for teaching. Most research on GTA professional development has not taken into account evidence of the social nature of learning that appears to be particularly relevant to GTAs. A CoP perspective for examining graduate student professional development is proposed that accounts for the perceived faculty leadership, social norms, orientation, and engagement that graduate students experience in their CoP.

The aim of this study is to investigate how variation in the four aspects of graduate students' experiences of CoPs may be related to each other and to student-centered teaching beliefs. Also of interest is exploration of whether relationships hold for different groups of graduate students (international vs. native; graduate students with prior teaching experience; graduate students with interest in teaching).

2.8.5 Research question. The research question for this study is: ***How are statistics graduate students' perceptions of their experiences in CoPs related to their***

beliefs about teaching statistics? To address this research question, relationships between the six constructs in the theoretical model in Figure 1 will be examined. Four of the constructs are designed to describe graduate students' perceptions of their experiences in CoPs (Engagement, Norms, Orientation, and Faculty Presence), and are based on aspects of members' experiences of CoPs found in the literature related to CoPs in other contexts. The other two constructs in the theoretical model describe graduate students' beliefs about teaching statistics (Prior Beliefs, Current Beliefs). Follow-up questions will include examination of whether the relationships are invariant under variables such as international student status, interest in teaching, and prior teaching experience.

Answers to this research question will lead toward a better understanding of how graduate student CoPs can be studied. Answers may also reveal how participation in different kinds of CoPs may be associated with graduate students' beliefs about teaching statistics. Factors that may affect graduate students' abilities to grow and develop as members of CoPs may be uncovered. The research may establish foundations for a greater understanding of how a CoP perspective might be able to provide sustainable professional development to graduate students in statistics departments.

Chapter 3

Methods

The literature in the previous section provided motivation for studying the professional development of GTAs for teaching statistics. The review also provided rationale for the use of a CoP perspective for studying graduate students' growth and development as teachers. A visual model was presented to identify possible associations between graduate students' perceptions of their participation in CoPs and their teaching beliefs (see Appendix A). The model has six core constructs: current beliefs about teaching statistics (Current Beliefs); prior beliefs about teaching statistics before entering current degree program (Prior Beliefs); participants' perceived level of engagement in their CoP (Engagement); participants' perceptions of the health of the norms of interaction in the CoP (Norms); the participants' perceptions of the CoP's orientation toward student-centered teaching (Orientation); and the participants' perceptions of faculty presence in the CoP (Faculty Presence).

This study seeks to explore the relationships between these core constructs, with a particular focus on how constructs may be associated with graduate students' beliefs about teaching statistics. This study also seeks to examine whether the relationships are invariant across different groups (e.g., international student status, prior teaching experience).

3.1 Overview of the Study

To explore the extent to which the graduate students' experiences of CoPs may be related to their beliefs about teaching statistics, an online survey was developed and administered to graduate students in statistics departments in the United States. The survey is given in Appendix B. The process for developing the survey had many stages, including a preliminary focus group, several drafts, think-aloud interviews, and a final pilot session.

The 70-item survey was administered to graduate students across the nation using Qualtrics (Qualtrics, Provo, UT, 2015), an online survey program. The survey administration began in February 2016, so that new graduate students in the 2015–2016 academic year would have a chance to engage with their CoPs for at least one semester or quarter by the time of the data collection. Participants were recruited in four waves, and this chapter includes a detailed account of the methods used to recruit participants and gather data. Typically, the survey took about 10 minutes to complete.

Once the data were collected and cleaned, confirmatory factor analysis of the results was used to create measures of the six core constructs in the proposed theoretical model. Path analysis was used to explore the proposed theoretical model, and variants of the model that were theoretically supported. Ordinary least squares regression was used for further analyses. Cross-validation techniques were used to select the final model and to explore the invariance of this model across different populations of interest.

3.2 Steps in Developing and Revising the Survey

The survey development process involved many steps including data collection from preliminary focus groups, instrument blueprint development, feedback from faculty and graduate students familiar with principles of survey design, think-aloud interviews, and a brief piloting phase. Each of these stages of the survey development process is described in detail in the following subsections.

3.2.1 Preliminary focus groups. Prior to the start of this study, two focus groups were conducted to obtain preliminary information that could be used to draft the survey. The focus groups, conducted at the 2015 *United States Conference on Teaching Statistics*, included eleven graduate students from seven institutions in four different US Regions (East, South, Midwest, and Southwest). Most participants represented departments of statistics, although a few were graduate students from other departments (e.g., mathematics education).

The focus group questions were designed to learn about graduate students' experiences interacting with other graduate students in their departments. It was hoped that answers to the questions would help identify ways of measuring different aspects of graduate students' CoPs. The initial questions used in the focus group are given in Appendix C. The interviewer followed up with additional questions if clarification was needed.

Information gathered from the focus groups helped reveal nuances in terminology. For example, some graduate students who have served as the primary instructor for courses do not consider themselves *graduate teaching assistants;* they use the term *graduate instructors*. This difference in nomenclature clued the researcher into the many different types of teaching assistantships available to graduate students, and helped prevent potential survey participants from mistakenly self-selecting themselves out of the target population. Also, there appear to be differences among graduate students as to who qualifies as "faculty" (e.g., full-time lecturers vs. tenure-track instructors).

These discoveries were used to appropriately define the target population and sampling frame, as well as to draft items for the survey instrument.

Other information gathered from the focus groups was used to develop some of the items in the survey blueprint. For example, the focus groups revealed to the researcher that one source of engagement in a CoP might be a shared office. Therefore the blueprint (and final survey) included items about whether participants spend time in a shared office with other graduate students.

3.2.2 Blueprint. Based on the information gathered from the focus groups and the literature reviewed in Chapter 2 of this paper, a blueprint for the survey was created. Appendix D offers the initial blueprint of the survey instrument, including first drafts of items for each of the topics to be measured. The blueprint topics include each of the six core constructs in the proposed theoretical model, as well as other characteristics that may be of interest based on the prior research. To avoid underfitting the model for potential latent variable analyses, at least four survey items were included to measure each of the constructs (Beaujean, 2014).

3.2.3 Item writing. For most of the constructs in the model, drafted items were either adapted from other instruments or developed based on information gathered from the preliminary focus groups. Examples of original items and their revised (final) versions are given in Appendix E. This subsection outlines the item development process and the sources from which items came for each of the six core constructs and the characteristics variables.

3.2.3.1 Constructs 1 & 2: Current and Prior Beliefs. Draft items for the outcome variable of the study, *Current Beliefs*, were taken from the *Graduate Students Statistics*

Teaching Inventory (GSSTI; Justice et al., in press), an instrument that is based on the GAISE recommendations for teaching introductory statistics (ASA, 2005). Some modifications were made to items from the GSSTI. For example, all the beliefs questions included specifications to respond according to one type of basic, face-to-face introductory course for non-statistics majors (see Appendix B, Introduction to Section 4: *Your Beliefs about Teaching and Learning*). This modification was made to avoid potential confounding due to the many different types of introductory statistics courses that can be offered (Justice et al., in press).

Parallel items were used for the common-cause variable, *Prior Beliefs*, which was used as a covariate in the final model. The use of retrospective measurement of prior beliefs is recommended to control for response-shift bias (Bray & Howard, 1980). As mentioned previously, the items regarding Prior Beliefs were presented in tandem with the Current Beliefs items so as to reduce cognitive load by requiring participants to read each stem only once.

3.2.3.2 Construct 3: Engagement. Items designed to measure the *Engagement* construct were developed based on data collected in the preliminary focus group sessions. For example, because focus group participants hypothesized that graduate students who share an office space are more apt to engage with one another about teaching topics, an item regarding the frequency of visiting a shared office was included. Items regarding the duration and frequency of meetings were also drafted as a result of data collected from the focus groups.

3.2.3.3 Construct 4: Norms of Interaction. Draft items for the *Norms of Interaction* construct were adapted from items used by Arbaugh et al. (2008). Their study used a survey to measure graduate students' perceptions of their participation in online communities of inquiry. When selecting items from Arbaugh's instrument, suggestions for studying CoPs offered by Little (2002) were used. For example, items that loaded highly on the "social presence" factor were given greater consideration to be included.

3.2.3.4 Construct 5: Orientation. Items from the GSSTI (Justice et al., in press) were used to draft items designed to measure the *Orientation* construct. The four original items selected were designed to measure the CoP's approach toward use of active learning methods, cooperative learning methods, alternative assessment methods, and an emphasis on conceptual ideas rather than mere knowledge of procedures. Appendix E offers an example of an item from the GSSTI and the corresponding item that was used in the final instrument.

3.2.3.5 Construct 6: Faculty Presence. For the *Faculty Presence* construct, one item was included to measure whether graduate students have experienced a teaching observation from a faculty member. This item was included because of empirical support for teaching observations as an effective form of professional development (e.g., Dalgaard, 1982). The rest of the items for this construct were adapted from items used by Arbaugh et al. (2008), whose survey study also addressed students' perceptions of Leadership Presence in online communities of inquiry. Appendix E offers an example of an item used by Arbaugh et al. (2008) and the final version.

3.2.3.6 Other characteristics. The blueprint also included items designed to measure characteristics of graduate students that are suggested by previous research to explain variance in graduate students' experiences in CoPs or their teaching beliefs. Where possible, items were adapted from the GSSTI 2. The GSSTI 2 is a modified version of the GSSTI (Justice et al, in press) that was developed by the author as part of a survey design class project in the fall of 2013. As part of the class project, items in the GSSTI 2 went through several rounds of peer and instructor revision, as well as some piloting with graduate students in applied statistics fields.

3.2.4 Instrument first draft. To create the first draft of the survey instrument, items were arranged according to principles of survey design (e.g., Dillman, Smyth, & Christian, 2009). Like-items were grouped together in order to reduce cognitive load as much as possible. Potentially sensitive student characteristics and topics (e.g., Norms of Interaction, international student status, gender, age, and desire to acquire a Ph.D. in the program) were reserved for the end of the instrument. Less sensitive questions, such as graduate students' frequency and duration of engagement in CoPs, were included in earlier sections.

Some items with Likert-type response options were modified so that the response options were estimated percentages on a 0–100 scale. This change was made primarily because continuous variables have advantages for analysis that are not offered to ordinal scales. The process of refining the items to the continuous scale had other advantages. For example, the change caused the researcher to notice some items that could use more clarification. An example of a formerly Likert-type response item that was changed to a continuous scale is given in Appendix E.

3.2.5 Initial Feedback. After the first draft was put together, feedback was sought from three members of an educational psychology department: a faculty member, a full-time lecturer, and an international graduate student, each of whom are familiar with principles of survey design via coursework or extensive survey-development experience.

Based on the feedback, some items were added and others were refined. For example, additional questions about Beliefs (current and prior) were added to complement the items regarding alignment with GAISE recommendations from previous instruments. These items were developed after reflective thought about the researcher's changes in teaching beliefs and practices over the course of her graduate school experience. In some cases, items were adapted from the Reformed Teaching Observation Protocol (Sawada et al., 2002).

Based on the initial feedback some of the items were refined. Items identified as measuring two constructs simultaneously (double-barreled items) were either split into two items or adjusted to measure just one construct (e.g., Bassili & Scott, 1996). Also, based on feedback from the international student, the wording of many items was simplified to become more accessible to students for whom English is not their first language.

3.2.6 Think-Aloud Sessions. Formal think-aloud sessions were conducted with five graduate students representing three research institutions in two regions of the United States (the Midwest and the Southeast). Participants were chosen from the researcher's previously-known contacts to represent a variety of institutions and backgrounds, and based on the researcher's belief that these students might be willing to spend about one hour to conduct the think-aloud session. Four females and one male participated. Two of the five think-aloud participants were international students.

Two participants were current members of the target population (graduate students in statistics departments). The other three participants were not current members of the target population, but either had experiences as graduate students in a related field, or had recently been members of the target population. The participants who had recently been members of the target population were asked to complete the think-aloud interview reflecting back as if they were still in their final year as a member of the target population.

Based on the think-aloud sessions, several additional changes were made to the instrument. The international students identified more language that was difficult to understand, and after some discussion appropriate substitutions were identified. For example, *supervisor* was changed to *overseer* because the former term was expressed as having harsh and overbearing connotations (Item 9, Appendix B). Also, some items were modified to improve clarity. For example, in items about teaching beliefs (Section 4, Appendix B), further specifications were offered about the nature of the hypothetical course (e.g., not calculus-based, about 35 students, with no additional recitation section).

Also based on results of the think-aloud sessions, additional prose was added to encourage participants to leave the beliefs items blank if they were unable to respond to them. This prose was added after one think-aloud participant chose to enter 0% when she was unable to estimate percentages reflecting her prior beliefs. Such a response would confound the results because 0% indicates a very extreme response rather than a neutral or NA-type response. It was preferred that graduate students who were unable to answer simply leave the items blank, even if it rendered their data unusable.

3.2.7 Final Pilot Session. Two additional graduate students in the target population were asked to take the survey as part of a brief pilot session. The graduate students were contacts that the researcher had met at previous statistics education conferences. They were chosen for the pilot session because they represented two large, mid-western institutions that were different from the institutions represented by think-aloud session participants. The pilot session participants were also selected to for this role because one was an international student and the other was a native student.

The pilot participants were asked to identify any issues with the survey such as broken links, typos, or questions that were unanswerable. The results of the pilot session offered minor typos and suggestions, but no major issues. The times to take the survey (15–20 minutes) for these participants were used to estimate the time in the initial advertisement letter. (After the first round of participant recruitment this estimate was shortened because the data suggested the reduced time interval was more appropriate for non-pilot participants).

3.3 Survey Instrument

The final version of the survey instrument contained 70 items organized into six sections. Table 3 gives the number of items in each section, while Appendix B gives the entire survey instrument. The order of the sections and items was guided by survey design principles (Dillman et al., 2009). For example, earlier sections were designed to include items that were straightforward and easy to answer; meanwhile the more difficult items or items regarding more personal information appeared later on.

Table 3.

GETS Inventory Items by Section

Section No.	Section Name	No. of Items
1	You and Your Graduate Program	10
2	Interactions With Others in Your Department	12
3	Faculty Support for Your Teaching	4
4	Your Beliefs About Teaching and Learning	18
5	The People Who Influence Your Teaching Beliefs	18
6	Getting to Know You	8
	Total	70

Note. The survey also included an initial item for participants to indicate consent for participation. A final open-ended item was included for any additional comments participants wish to provide. Items for participants to enter the drawing for the prize are not reflected in this table.

Items in Section 1 gather demographic information about participants. Section 1 also contains questions to ensure that respondents are members of the target population. For example, an initial item asks whether respondents are current graduate students. Respondents who did not identify as current graduate students were immediately thanked and routed out of the survey.

Sections 2–3 are designed to measure two of the six core constructs in the theoretical model. Section 2, *Interactions With Others In Your Department*, contains items designed to measure Engagement. Faculty Presence, another core construct from the theoretical model, is represented by three items in Section 2 and four items in Section 3, which is entitled, *Faculty Support for Your Teaching*.

Section 4, *Your Beliefs About Teaching*, includes 18 items designed to measure Prior Beliefs and Current Beliefs. To reduce the cognitive load required to complete the survey, each item about Prior Beliefs was arranged to follow immediately after the corresponding question about Current Beliefs. This organization was chosen so that participants did not need to read the stem for parallel items twice. Items regarding Current Beliefs were included first to anchor participant responses according to the (more important) outcome variable of the study, Current Beliefs, rather than a covariate, Prior Beliefs.

Section 5, *The People Who Influence Your Teaching Beliefs*, includes 18 items designed to measure the final two constructs of the theoretical model, Norms and Orientation. To measure Norms, respondents were asked about interactions with the two members of their department with whom they believed their relationship to be most influential. To measure Orientation, participants were asked about the teaching practices of the person in their department who they perceive to have the largest influence on their current beliefs.

Section 6, *Getting to Know You*, has items designed to gather more demographic information about participants and characteristics of interest that were considered too personal to include in Section 1. For example items designed to gather information about international student status and years of experience teaching prior to becoming a graduate student were included in this section.

3.4 Target Population

The target population for the study was defined to be all graduate students in statistics departments in the United States. The target population seeks to capture current students who may become the future statistics professorate. To that end, the target population was viewed as students in statistics departments that offer doctoral degrees.

Graduate students who consider themselves to be master's students but who are in departments that offer doctoral degrees were included in the target population, because many of them may go on to become doctoral students. Also these students may have similar experiences as those who consider themselves to be doctoral students in the same programs. Students who are in departments that offer only masters degrees were not included in the target population, because it is plausible that their experiences are fundamentally different because they may be more focused on preparation for teaching-related positions.

Although the main focus of the study is statistics graduate students who participate in CoPs and have experiences teaching or assisting with teaching statistics (GTAs), all statistics graduate students were included in the target population. There are two reasons for using the larger population. First, including non-GTAs allowed for exploring the extent to which graduate students not involved in teaching still paticipate in CoPs related to teaching. Secondly, all graduate students were included because it was found in the preliminary focus groups that the term *GTA* is not consistent among departments. Some students did not consider themselves GTAs because they were the instructor of record for a course so identified as graduate instructors rather than graduate teaching assistants. On the other end of the spectrum, some students who have assisted with courses did not consider themselves GTAs because they were *not* the instructor of record or did not have teaching responsibilities that they considered sufficient to identify as a GTA. To prevent participants from self-selecting themselves out of the pool of potential participants, it was decided to include all graduate students in the target population.

Although the original target population focused on graduate students in statistics departments, graduate students in biostatistics departments were included as well. Biostatistics students were included to allow for exploration of whether models are invariant across the type of statistics department (statistics vs. biostatistics).

3.5 Participant Recruitment

Participant recruitment occurred in three waves. First, the Executive Director of the ASA posted a solicitation for participation in the Caucus of Academic Reps Weekly Digest, an electronic newsletter for the ASA's community of academics interested in statistics. The solicitation, which is offered in Appendix F, was posted on February 5, 2016. Faculty who are members of the Caucus were asked to forward the survey invitation to their graduate students. This method of data collection appeared to gather no participants.

Secondly, the executive director of the ASA sent two follow-up e-mails directly to members of the Caucus of Academic Representatives. The initial e-mail was sent on the evening of February 16, 2016. The follow-up was sent on February 24. The e-mails again asked department chairs to forward to their graduate students an invitation containing a link to the survey. The e-mails are included in Appendices G and H. At least 129 (not all useable) participants were recruited using this method.

In the third wave of participant recruitment, e-mails were sent to contacts (student and faculty) in statistics departments who had previous relationships with the University of Minnesota Statistics Education program or with the researcher. If no respondents from these institutions appeared, a follow-up e-mail was sent about one week later. Example e-mails and follow-up e-mails are given in Appendices I and J.

The survey closed on 15 March 2016 at 11:59 PM. There were 387 participants, but many were not complete or indicated they were not in the target population. There were 248 participants from statistics and biostatistics department who finished the survey. Many of these cases were not useable due to missing responses. For the final chosen model, 218 cases were useable.

All invitations to participate in the research followed principles of social exchange given by Dillman et al. (2009). That is, they emphasized benefits for participation (e.g., contributing to research on GTA professional development, being entered into a lottery for one of five Amazon.com gift cards), developing trust (e.g., naming the purpose of the study explicitly), and minimizing costs (e.g., the survey takes 10–15 minutes to complete). To bolster response rates, appeals were made to participate based on the fact that data would be used for a dissertation study. It was hoped that graduate students may empathize with another graduate student in need of data for her degree completion.

3.6 Chapter Summary

To explore the extent to which graduate students' experiences in CoPs are related to their beliefs about teaching statistics, the GETS Inventory was developed. Items in the instrument were modified based on feedback from think-aloud interviews with graduate students who are (or were in the past) members of the target population. The final survey instrument included 70 items designed to measure the six core constructs of the study and participant demographic information. Invitations were sent to faculty who are in the Caucus of Academic Representatives in the American Statistical Association. Invitations were also sent to faculty contacts having connections with the researcher for this study. Faculty were asked to forward the invitation to graduate students in their departments. Chapter 4 offers the results of the survey and analysis.

Chapter 4

Results

To explore the relationships between graduate students' participation in CoPs and the extent to which their beliefs about teaching statistics are student-centered, a survey was created and administered to statistics graduate students across the United States. This chapter describes how the results were used to define measures for six core constructs: graduate students' current beliefs about teaching statistics (Current Beliefs); graduate students' beliefs about teaching statistics held just prior to entering their current degree programs (Prior Beliefs); the extent to which graduate students engage in CoPs related to teaching (Engagement); the extent to which the interactions in the CoP are healthy (Norms); the extent to which the CoP holds teaching beliefs that are student-centered (Orientation); and the extent to which participants perceive faculty to be involved in the CoP (Faculty Presence). In addition to models relating the core constructs, models were examined that included variables describing characteristics of graduate students (e.g., year in program, international student status).

Initial data cleaning involved removal of participants who did not finish the survey, participants who did not indicate that they are currently graduate students, or participants whose time stamps indicated that they did not give any thought to the responses. Participants from math education or educational psychology departments were removed from the sample because it is believed that their experiences would be focused more on teaching and education than those in the population of interest. Also, students who listed other departments such as computer science, parks & recreation, biomedical informatics, and engineering were removed for the same reason. Initially, the 8 cases in mathematics departments were included on the grounds that many statistics graduate students may be housed in Mathematics-Statistics Departments. However, this decision was reconsidered when the model results were quite different when the mathematics department students (n=8) were removed versus when they were included. Therefore the results presented in this chapter represent participants who indicated their degree programs are housed in statistics or biostatistics departments.

4.1 Description of the Sample

All exploratory data analysis and visualization were conducted using the open-source computing program, R version 3.1.0 (R Development Core Team, 2013), and ggplot2 (Wickham, 2009). We begin with an overview of characteristics of the participants in the survey. Table 4 gives summaries of the characteristics of the sample.

Table 4.

Description of the Sample

Characteristic	Used in Model (N=218)		All Responses (N=245)	
	N	%	N	%
Gender				
Female	111	51	130	53
Male	106	49	113	46
Does not identify as F or M	1	0	2	1
Department type				
Statistics	152	70	173	71
Biostatistics	66	30	72	29
Expecting to earn PhD in current department				
Yes	168	77	188	77
No	34	16	37	15
Undecided	16	7	19	8
International student				
Yes	55	25	63	26
No	163	75	181	74
Prior experience teaching K–12				
Yes	23	11	24	10
No	195	89	221	90
Prior experience teaching college				
Yes	21	10	21	9
No	197	90	224	91
Expect to teach as part of career				
Yes	103	47	109	44
No	115	53	136	56

Note. The larger set described includes the 245 participants in statistics and biostatistics departments who completed the survey, and indicated they were currently graduate students in institutions that offer doctoral degrees. The smaller set excludes the 27 participants who were not able to report enough information about prior beliefs to have scores imputed and used in the final model.

Initially, the useable data included 245 participants. However, missing data were an issue for most of the base models that were considered. In particular, prior beliefs were difficult for participants to identify; 27 participants were not able to report enough information about prior beliefs to have scores imputed and used in the final models. Unless otherwise noted, the prose that follows describes the set of responses useable in the final base model (N=218).

While the sample results indicate gender is mostly an even split between males and female participants (with very few participants identifying as neither male nor female), other characteristics are not as evenly distributed. The sample has over two times more participants from statistics departments than biostatistics departments. Also, the sample is weighted toward students who expect to complete a doctoral degree in their current programs (as opposed to completing a master of science degree, or neither). One-quarter of participants are international students. Prior to entering their degree programs, about one-tenth of participants had experience teaching at primary and secondary levels, and a similar fraction acquired prior teaching experience at the college level. Three percent of participants had prior teaching experience at both levels. About half of participants indicated that they expect to teach as part of their career. The final useable data set represented 37 institutions from all major regions of the United States. A list of represented institutions is given in Appendix K.

To help get a sense of characteristics of participants, age and year in program were also collected for each participant. Five number summaries for these variables are given in Table 5. As one might imagine for graduate students, the majority of participants were age 22–30. Less than 15 participants indicated they were older than 30 years of age.

Table 5.

Quantitative Characteristics of the Participants Whose Responses Were Used in the Final Model

Characteristic	*N*	*M*	*SD*	*Min*	*Q1*	*Med*	*Q3*	*Max*
Age	216	26.8	5	21	24	26	28	54
Years in program	218	2.8	2.5	0	1	2	4	32

Note. N=218. Participants who did not indicate that they were currently graduate students were removed from the sample.

Information was also collected regarding participants' experiences with teaching and research responsibilities in their current degree programs. About 80% reported having been hired for at least one of the teaching or assistant-related position offered in the survey (e.g., grading papers, facilitating discussion or lab sessions, teaching a course). Roughly 30% reported having served as a primary instructor for a course. Around 10% have also served as supervisor of other GTAs regarding teaching responsibilities. About 60% reported prior experience in a research-assistant position.

The following subsections offer descriptive statistics regarding participants' perceptions of their beliefs about teaching statistics, engagement in CoPs, orientations of their CoPs toward student-centered teaching, norms of interaction in their CoPs, and faculty presence in their CoPs. Only participants whose data were used in the final model (*N*=218) are included.

4.1.1 Beliefs about teaching statistics. Information was also collected to learn about the beliefs that participants hold about how statistics should be taught. There were nine items about participants' current beliefs about teaching statistics and nine items about participants' beliefs prior to entering their current degree programs (Items 27–44, Appendix B). Table 6 gives the results regarding participants' current and prior beliefs. For all items participants chose a percentage (i.e. 0–100) that indicated their beliefs about how often each topic should be used.

Table 6.

Current and Prior Beliefs About Teaching Statistics

	Current		Prior		ΔCurrent–Prior	
Topic	M	SD	M	SD	M	SD
Percentage of Class …						
Time used on Instructor Lectures*	59	23	71	22	–12	19
Time used for Group Activities	23	19	15	15	8	16
Time Instructor Explains Misconceptions*	21	16	20	18	1	11
Sessions that Instructor Delivers Content first*	52	29	59	30	–8	18
Percentage of Assessment(s)…						
Similar to Previous Examples*	70	21	72	23	–2	16
Requiring Explanations	46	24	37	23	9	18
Completed in Small Groups	22	17	18	16	4	16
(Homework) Uses Procedures and Formulas*	42	23	50	25	–8	21
Percentage of Inference Taught Using Simulation	31	23	17	19	14	20

Note. Based on *N*=218 participants used to compute the final model.
*Would be reverse-coded to indicate student-centered teaching practices

Also of interest are participants' perceived changes in beliefs since the time they entered their current degree programs. The final two columns in Table 6 display the average changes in beliefs (Current – Prior). After reverse coding when appropriate, the table indicates that for most topics, students' beliefs have typically progressed toward more student-centered teaching beliefs since the time they entered their degree programs. The only topic that does not indicate change in the student-centered direction, on average, is for the percentage of class time that the instructor uses to explain misconceptions.

Figure 2 illustrates the distributions of the change (current minus prior) for each of the nine items regarding beliefs about teaching statistics. Responses are reverse coded when appropriate so that positive values represent changes toward more student-centered beliefs. For most topics, the first quartile is at or above 0, indicating that roughly three-quarters of the participants have changed either not at all, or toward more teacher-centered beliefs. For the middle 50% of participants, changes tend to appear to be between 0–20 percentage points in the student-centered direction. There appear to be more outliers on the right side of the plots; that is, more often participants perceive themselves to have drastic changes in the student-centered direction than the teacher-centered direction. Participants' year in the program was not a significant predictor of the magnitude of their change in beliefs for any of the nine beliefs items.

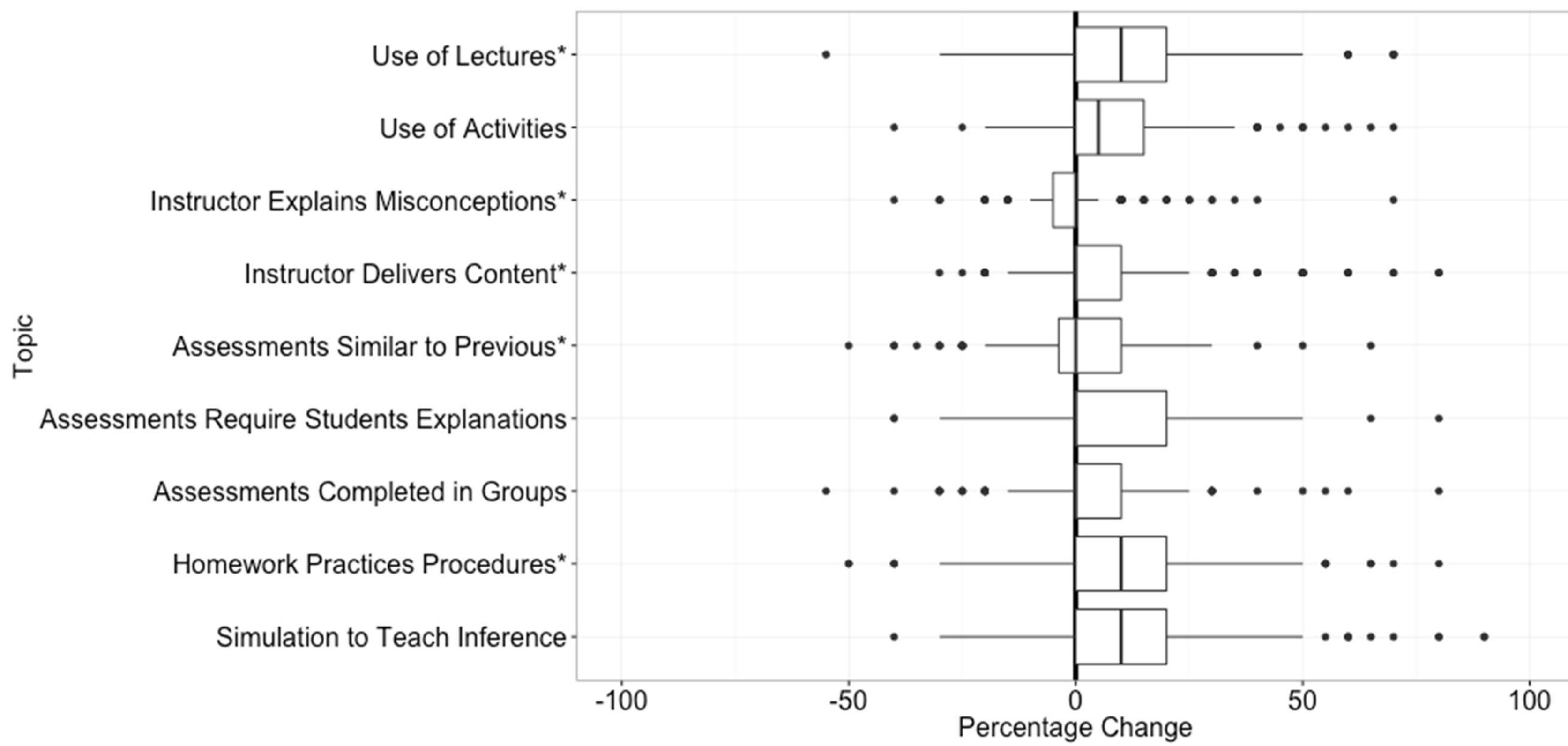


Figure 2. Changes in Beliefs Since Entering Current Degree Program. Topics with an asterisk have been reverse-coded so that positive changes indicate movement toward student-centered teaching beliefs.

The correlations between items were also examined. Table 7 gives the pairwise-complete intercorrelations for the current beliefs items. Table 8 gives the pairwise-complete intercorrelations for the prior beliefs items. Items were reverse coded, when appropriate.

It is clear from the table that some item pairs have weak or negative, while others had strong correlations. For both current and prior beliefs, Topic 3 (instructor explains misconceptions) and Topic 5 (novel exam problems) are negatively correlated with many of the other items, and do not appear to be measuring the same construct as the others. Topics that appear to have strong positive correlations with each other are use of lectures, the use of activities, the use of group assessments, and the use of simulation to teach inference.

Table 7.

Pairwise-Complete Intercorrelations of Topics Regarding Current Beliefs about Teaching Statistics

Topic (Item Number)	1	2	3	4	5	6	7	8	9
1. Use of Lectures* (41)	—								
2. Use of Activities (37)	.57	—							
3. Instructor Explains Misconceptions* (39)	–.06	–.16	—						
4. Instructor Delivers Content* (43)	.44	.18	.20	—					
5. Exam Problems Similar* (29)	.11	–.01	–.08	.17	—				
6. Exams Require Explanations (31)	.18	.33	–.12	–.02	–.11	—			
7. Assessments in Groups (33)	.21	.35	–.2	.00	.03	.28	—		
8. HW Practicing Procedures* (27)	.22	.06	.12	.11	.15	.11	–.07	—	
9. Simulation-Based Inference (35)	.26	.30	–.29	.03	.16	.38	.34	.11	—

Note. Based on pairwise complete observations of *N*=218 participants used to compute the final model.
*Reverse coded.

Table 8.

Pairwise-Complete Intercorrelations for Topics Regarding Prior Beliefs about Teaching Statistics

Topic (Item Number)	1	2	3	4	5	6	7	8	9
1. Use of Lectures* (42)	—								
2. Use of Activities (38)	.50	—							
3. Instructor Explains Misconceptions* (40)	–.06	–.25	—						
4. Instructor Delivers Content* (43)	.45	.24	.16	—					
5. Exam Problems Similar* (30)	.08	–.03	–.02	.12	—				
6. Exams Require Explanations (32)	.12	.22	–.28	–.02	–.11	—			
7. Assessments in Groups (33)	.13	.46	–.28	.03	–.10	.29	—		
8. HW Practicing Procedures* (28)	.14	–.01	.08	.16	.26	.00	–.09	—	
9. Simulation-Based Inference (36)	.15	.29	–.22	.05	.07	.24	.31	.11	—

Note. Based on pairwise complete observations of N=218 participants used to compute the final model.
*Reverse coded.

4.1.2 Participation in CoPs. Information was also collected to try to measure the extent to which respondents participate in CoPs. Perhaps the easiest way to detect participation in CoPs was attendance of required meetings regarding teaching. About 75 percent of participants attend required meetings with graduate students, faculty, or staff to discuss topics related to teaching or teaching responsibilities. Of these participants, about half indicated that meetings are weekly. The next highest proportion, one-third, indicated required meetings occur fewer than once per month. Participants have not been required to attend the meetings for very many years. Over half of the participants required to attend meetings indicated that the meetings have lasted for one year or less. Three-quarters indicated they have participated in required meetings for two years or less.

Another line of questions asked about participation in voluntary meetings and discussions regarding teaching (Items 17–19, Appendix B). As with required meetings, about 75 percent of participants indicated that they have engaged in voluntary meetings. Most of these participants (about 80 percent) were the same participants who attend required meetings. The frequency of voluntary meetings varied. The distribution was fairly uniform across five categories: fewer than once per month, monthly, two to three times per month, weekly, and two or more times per week. About half of those who participate in voluntary meetings have done so for no more than one year. About 30 percent have participated for one to three years.

There appears to be a relationship between participation in meetings regarding teaching topics and the extent to which graduate students have responsibilities related to teaching. Table 9 gives the frequencies of participation in voluntary and required meetings, grouped by experience as an instructor of record, assistant to the instructor of record, or no teaching responsibilities. The percentage of participants who attend required meetings progressively decreases as teaching responsibilities decrease (91, 86, and 16 percent, respectively). Engagement in voluntary meetings follows the same pattern, but with a less drastic drop in the last category (88, 74, and 66 percent, respectively).

Table 9.

Counts (and Percentages, Calculated Within Level of Teaching Responsibilities) of Participation in Meetings about Teaching, Grouped by Highest Experience Level of Teaching Responsibility.

	Required Meetings		
Voluntary Meetings	No	Yes	Total
Instructor of Record			
No	1 (1)	7 (10)	8 (12)
Yes	5 (7)	56 (81)	61 (88)
Total	6 (9)	63 (91)	69 (100)
Any Teaching Responsibilities			
No	4 (4)	25 (23)	29 (26)
Yes	12 (11)	70 (63)	82 (74)
Total	16 (14)	95 (86)	111 (100)
No Teaching Responsibilities			
No	12 (32)	1 (3)	13 (34)
Yes	20 (53)	5 (13)	25 (66)
Total	32 (84)	6 (16)	38 (100)

Note. N=218. Teaching responsibilities are for current degree programs. Assistant to the instructor includes tasks such as grading assignments or holding office hours. Voluntary meetings include unplanned discussions and other informal interactions regarding teaching topics.

To learn about the CoPs they experience, participants were asked how many graduate students and faculty they felt comfortable approaching to discuss teaching topics. Both distributions were skewed to the right with several modes. Modes for number of approachable faculty were at 2–3, 5, and 10 faculty members, while modes for approachable graduate students were slightly higher values: dominant modes at 5 and 10, with smaller modes at 15, and 20. Table 10 gives the five number summaries of the results of these two items. Only three participants indicated that they do not feel comfortable approaching any other graduate students to discuss teaching-related topics. Three participants indicated zero for the analogous question regarding approachable faculty. Only one participant indicated that they felt comfortable approaching no faculty or graduate students regarding teaching topics.

Table 10.

Five Number Summaries of Variables Collected to Indicate Level of Participation in Communities of Practice

Variable	*M*	*SD*	*Min*	*Q1*	*Med*	*Q3*	*Max*
No. of Faculty Approachable	7	10	0	3	5	7	90
No. of students approachable	16	18	0	5	10	20	100
Shared office visits / week	4	1.6	0	3	5	5	7

Note. Faculty and Students approachable is based on N=218 participants used to compute the final model. Visits to shared office per week is based on participants used to compute the final model who indicated that they do have a shared office (n=185).

The final items used to learn about participants' CoPs were whether they have a shared office, and the number of times they visit their office. Most participants (85 percent) indicated that they have a shared office, and of these participants, the most often number of visits per week was five (43 percent). Table 10 gives the five number summary of number of visits per week for those who indicated they do have a shared office (*n*=185).

Some of the items designed to measure participation in CoPs appear to be related. Table 11 gives the polychoric, polyserial, and Pearson correlation matrix for the five variables: number of office visits per week, number of approachable faculty, number of approachable graduate students, estimated number of required meetings per week, and estimated number of voluntary meetings per week.

Table 11.

Intercorrelations for Variables Designed to Measure Engagement in Communities of Practice

Variable	1	2	3	4	5
1. Number of Approachable Faculty	—				
2. Number of Approachable Students	.71	—			
3. Visits to shared office per week	.05	.05	—		
4. Frequency of required meetings[a]	.12	.12	.33	—	
5. Frequency of voluntary meetings[a]	–.05	–.01	.37	.18	—

Note. Based on *N*=218 participants used to compute the final model. Participants with no shared office were included as 0 visits per week. Zeros were also imputed for no voluntary and no required meetings. Significance tests were not included because tests of bivariate normality were highly significant. Polyserial or Polychoric correlations were calculated instead of Pearson correlations whenever appropriate for categorical variables.
[a]. Required and voluntary meetings were treated as ordered factors with the categories: fewer than once per month; monthly, 2–3 times per month, weekly (or more). The voluntary meetings had one additional category: 2–3 times per week (or more).

The number of approachable faculty had a strong positive correlation with the number of approachable students. Although there were a handful of outliers driving the strength of this correlation coefficient, there clearly was a positive relationship even when outliers were removed. Frequency of required meetings was positively correlated with all four of the other variables. The frequency of voluntary meetings was positively correlated with the frequency of visits to a shared office. It is interesting to note that the frequency of voluntary meetings was not positively correlated with either of the number of approachable faculty or the number of approachable graduate students. Visual examination of the plots suggests that outliers in number of approachable people do not seem to be the reason for these weak relationships.

4.1.3 Norms of Interaction in CoP. The survey included items designed to gather information about the health of interactions in CoPs. As a proxy for the community's norms of interaction, participants were asked to identify their two most influential department members and respond to questions about their interactions with these two members.

Table 12 gives summaries of the responses. There was not much variation; most responses indicated healthy interactions. It is notable that percentages were slightly higher for the person chosen as most influential compared to the second most influential person.

Table 12.

Norms of Interaction with the Two Department Members Chosen as the Most Influential Upon Participants' Teaching Beliefs

Topic	Most Influential *N*	Most Influential % Agree	2nd Most Influential *N*	2nd Most Influential % Agree
Comfortable Engaging in Discussion	217	96	216	89
Acknowledges Point of View	216	93	216	88
Can Respectfully Disagree	216	92	215	85
Cares About Teaching Quality	216	91	216	84
I Admire Person X As a Teacher	217	95	215	83

Note. Results Based on Items 47-56 (see Appendix B).

4.1.4 CoP Orientation toward student-centered teaching. Information was also collected to try to get a sense of the extent to which CoPs are oriented in favor of student-centered teaching. As a proxy for the community's beliefs, participants were asked to identify their most influential department member and answer questions about their teaching practices. Items 57–62 (see Appendix B) were used for this construct. To reduce cognitive load, items were written to be dichotomous.

Table 13 gives summaries of the six items designed to measure this construct. Items for which most primary influences are student-centered include the use of a variety of modes of communication (e.g., clickers, oral presentations; 84%) and the use of small-group activities (61%). Items for which the primary influencers appear to be largely teacher-centered are the use of lectures to deliver content (72%) and requiring the frequent practice of procedures using formulas (72%). Responses were somewhat split regarding the use of individual assessments completed in class (58%) and the use of simulation to teach inference (52%).

Table 13.

Orientation Toward Student-Centered Teaching of the Department Member Who is Perceived to be Most Influential Upon Teaching Beliefs

Topic	N	% Yes	% No
Content is presented mostly through lectures*	217	72	27
Content is presented mostly through small-group activities	216	61	38
Frequently requires students to practice procedures using formulas*	217	72	27
Uses simulation as the primary tool to teach inference	216	52	47
Assessments primarily in-class, individually completed*	217	58	41
Students communicate using a variety of means and media	217	84	15

Note. Based on Items 57–62 (see Appendix B). Participants were asked to respond based on introductory course to a class of about 35 students, not calculus-based, serving as a general university requirement (i.e., the students' majors do not have a statistics requirement). Percentages may not add to 100 because of rounding. Responses are based on N=217 participants; except in the case of one participant who was not able to answer items about the 2nd and 4th topics (N=216).
*Would be reverse-coded to represent student-centered teaching, according to theory.

Table 14 gives the polychoric correlations of the dichotomous items regarding the teaching practices of the influential department member. When appropriate, responses have been reverse coded so that positive correlations indicate teaching practices in the same direction. Nearly all of the correlations were positive, and many were quite large. The only two correlations that did not appear to match theory was the correlation between use of lectures (reverse coded) and simulation, and the correlation between frequent time spent practicing procedures (reverse coded) and students having a variety of means for communicating their ideas.

Table 14.

Polychoric Intercorrelations of Topics Designed to Measure CoP Orientation Toward Student-Centered Teaching

Topic	1	2	3	4	5	6
1. Use of Lectures*	—					
2. Use of Activities	.32	—				
3. Frequent Practice of Formulas and Procedures*	.45	.14	—			
4. Use of Simulation to Teach Inference	–.15	.43	.07	—		
5. Assessments In-class, Individually Completed*	.61	.29	.53	.08	—	
6. Students Communicate in a Variety of Modes	.35	.52	.01	.38	.11	—

Note. Based on *N*=217, pairwise complete observations, except for when to do with topics 2 and 4 (*N*=216).
*Reverse coded.

4.1.5 Faculty presence. Items 23–26 in the GETS inventory were designed to gather information about the extent to which faculty provide support in ways suggested by the literature to be important to graduate students involved with teaching responsibilities (e.g., Green, 2010). These items were only asked of those students who indicated that they have had some form of teaching responsibilities in their current degree programs.

Responses indicated fairly strong faculty presence for three out of the four items. Just over 90 percent of participants indicated that faculty clearly communicated required tasks for their teaching responsibilities. Just over 80 percent indicated that faculty identified topics that should be emphasized in the courses they teach or assist. Just under 70% of participants indicated that faculty facilitated productive conversations. Finally, just less than half indicated that faculty have observed them performing their teaching responsibilities at least once in their degree programs and provided feedback designed to help them improve.

Another set of items used to describe faculty presence is the extent to which faculty attend required and voluntary meetings that graduate students attend regarding teaching. The results are given in Table 15. For required meetings results were bimodal, with one mode at 0 times per month and the other at 4 times per month (weekly required meetings with faculty). As for voluntary meetings, there is one clear mode at 0 meetings attended by faculty, and the results tapered off immediately to the right, with few participants (~15%) indicating that faculty attended voluntary meetings at least once a month, and very few participants (less than 1%) indicating that faculty attended voluntary meetings weekly or more.

Table 15.

Summaries of Faculty Attendance of Voluntary and Required Meetings

	Number of Meetings (Per Month) Attended By Faculty						
Type of Meeting	M	SD	Min	Q1	Med	Q3	Max
Required	1.6	1.7	0	0	.5	4	4
Voluntary	.4	.7	0	0	.2	.5	4

Note. Based on *N*=218 participants used to compute the final model.

One unexpected result informs the extent to which faculty have a strong presence was the frequent choice of faculty as the most influential person on graduate students' beliefs about teaching statistics. A faculty member was perceived to be the most influential person on graduate students' beliefs by 83 percent of participants. Furthermore, nearly half of participants chose faculty members as the second most influential person as well.

4.2 Calculation of Measures

Participant responses were used to create measures of each of the six core constructs of the study: Current beliefs, Prior beliefs, Engagement, Norms, Orientation, and Faculty Presence. For each construct confirmatory factor analyses (CFA) was conducted to compare several candidate models. Responses to items were reverse-coded when theory suggested it was appropriate. The correlation matrices were calculated using pairwise complete observations. When appropriate for dichotomous variables and ordered factors, polyserial and polychoric correlations were used (Holgado-Tello, Chacón-Moscoso, Barbero-García, & Vila-Abad, 2010). These were calculated using the polycor package in R (Fox, 2016). Models were examined for coefficients that matched theory. Among the remaining candidate models the AICc was used, primarily, to compare them. The TLI and RMSEA were also examined for adequate fit. CFA coefficients and fit measures were calculated using the lavaan package in R (Rosseel, 2012).

4.2.1 Measure of current beliefs. To create a measure of the extent to which participant current beliefs are student-centered, responses were used from Section 4 of the survey instrument: Your Beliefs about Teaching. These items asked participants to offer a percentage (0–100) based on their beliefs about how an introductory statistics course should be taught. The items can be found in Appendix B.

Three different strategies were considered for translating responses into data to be input into candidate models for the measure. For one strategy, the original continuous values (0–100) were used. In another strategy, trichotomous values were assigned to each participant for each item, based on cut-points of the first and third quartiles of responses. In a third strategy, dichotomous scores were assigned based on cut-points at the median. When confirmatory factor analysis indicated better measures of model fit using the dichotomous scores, this (more simple) cut-point method was used for the remaining analyses.

There were three stages in the process for arriving at the median as the cut-points for the dichotomous scores. Cut points were initially determined by the researcher's theory of what it means to be student-centered. However these cut points were discarded upon examination of the distributions of responses; the researcher's ideas of student-centered teaching were too ambitious and did not allow adequate variation among responses (most participants would fail to achieve a score other than 0). Next, the distributions were examined for natural cut-points (e.g., between two modes) that allowed for adequate variation among respondents. Once it was discovered that the natural cut point was typically either equal to or very near the median of responses, the median response was used instead. Responses exactly at the median were determined to be student-centered (given a score of 1 instead of 0 after reverse coding was completed, if applicable).

After items were appropriately reverse-coded, two types of correlations between items were examined to create candidate models for the beliefs measure. First, polychoric correlations of the dichotomous scores were examined to identify items that may be appropriate for removal from the belief measure. Items with negative correlations were flagged for removal from candidate models. Secondly, correlations from the original (continuous) items were also used to flag items that may be removed from candidate models. For all pairs that had negative correlations, it was considered which item might have stronger theoretical justification for inclusion. Based on theory and the correlation coefficients between items, six candidate models were considered for the measure of graduate students' current beliefs. The six candidate models that were examined and the results of confirmatory factor analysis are given in Table 16.

Table 16

Confirmatory Factor Analysis Results for Models of the Current Beliefs Measure

	Coefficient Estimates Item Number									Fit Measures		
Model	35	31	33	41*	37	43*	27*	29*	39*	AICc	RMSEA	TLI
1	.28	.47	.32	.65	.54	.37	.17	.19	–.20	5930	.069	.723
2	.27	.45	.30	.67	.52	.40	.18	.19	—	5263	.081	.693
3	.26	.47	.31	.65	.54	.39	.17	—	—	4596	.085	.729
4	.27	.48	.34	.62	.57	.36	—	—	—	3927	.105	.691
5	**.33**	**.53**	**.39**	**.50**	**.61**	—	—	—	—	**3270**	**.032**	**.973**
6	—	.41	.26	.74	.49	.44	—	—	—	3264	.115	.719

Note. Bold indicates the chosen model. Item numbers correspond to items in the GETS Inventory (see Appendix B). Briefly: Item 35 represents the use of simulation for teaching inference; Item 31 represents the requiring of explanations using written words on exam questions; Item 33 represents assessments completed in groups; Items 41 and 37 involve the use of lectures and activities to deliver content, respectively; Item 43 involves the instructor explaining content before students use it; Item 27 involves the practice of procedures using formulae; Item 29 represents the use of novel assessment questions; and Item 39 involves the instructor explaining misconceptions to students.
*Reverse-coded

Despite the fact that Model 6 had lower AICc, Model 5 was chosen as the final model for measuring the current beliefs construct. Model 5 was the only one with RMSEA and TLI values that meet conventional standards for good model fit (e.g., Hu & Bentler, 1999; <.06 and >.9, respectively). Also, Model 5 is supported by theory in that it loads most heavily on the three items to do with using activities, explaining reasoning and using less lecture. These are more commonly-held indicators of student-centered teaching than the other two items regarding use of simulation methods and group tests (e.g., Kember, 1997). Using Model 5 means that participants' scores for the current beliefs measure were based on reported uses of simulations for inference, assessments that require explanations using words, group tests, lecture to introduce content (reverse coded), and activities.

For participants missing only one out of the five items needed to calculate a score (n=2), a score was imputed based on adjusted weights from the other four items for which the participant was able to respond. Responses from participants missing two or more of the five items were deemed unusable (n=11).

After scores were imputed for appropriate participants, scores were centered and scaled to have mean equal to 0 and standard deviation equal to one. The distribution of scores is given in Figure 3. Scores on to the right indicate beliefs that are more student-centered. The distribution is skewed to the left.

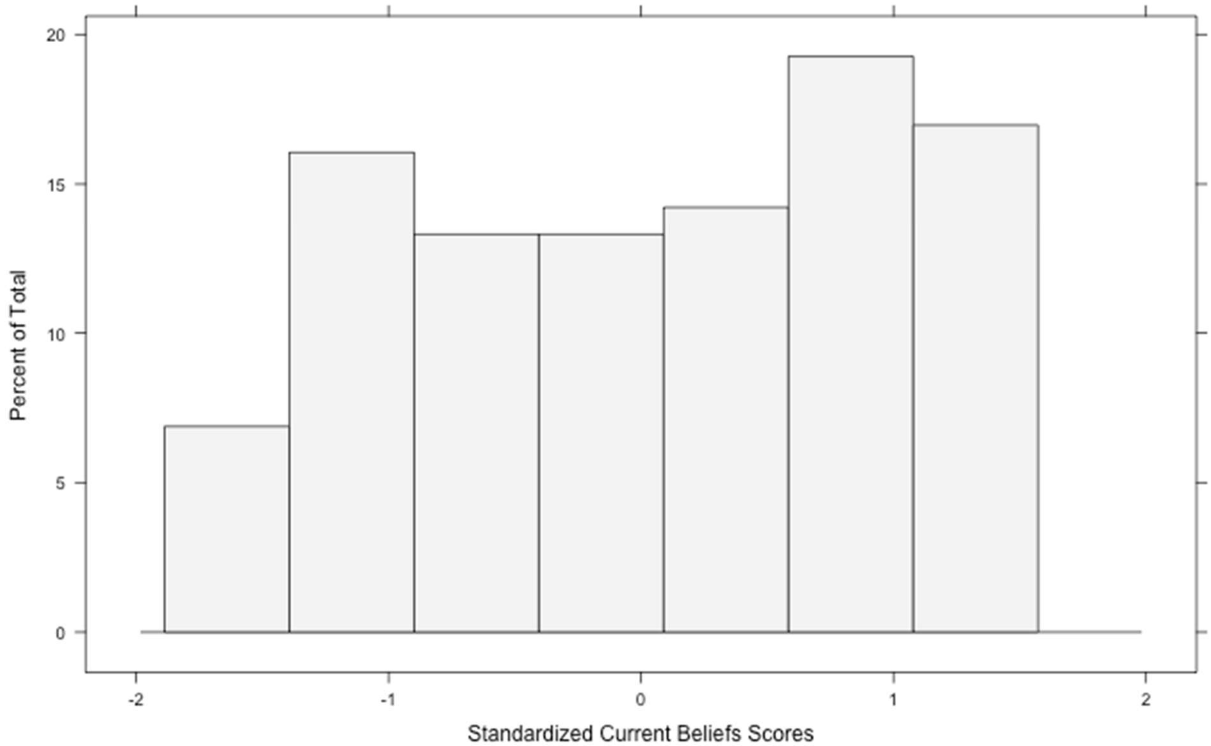


Figure 3. Distribution of scores for the measure of Current Beliefs (*N*=218). Scores were weighted based on CFA coefficients, and standardized after imputation of appropriate missing values.

4.2.2 Measure of prior beliefs. To create a measure of the extent to which participants' prior beliefs were student-centered, two candidate models were considered. The two candidate models use the same set of topics from the survey used to calculate the Current Beliefs measure. However the two models differ in how the items are assumed to load on the Prior Beliefs construct. One model used the same coefficients as were used to calculate scores for the measure Current Beliefs, while the other candidate model allowed the coefficients to vary. Both models used the same cut-points as were used for the Current Beliefs measure to create dichotomous scores. The coefficient estimates and fit measures for the two candidate models for the Prior Beliefs measure are given in Table 17.

Table 17

Coefficient Estimates and Fit Measures for Models of Prior Beliefs Measure

	Coefficient Estimates (Item Number)					Fit Measures		
Model	36	32	34	42*	38	AICc	RMSEA	TLI
1	.33	.53	.39	.50	.61	3126	.066	.856
2	**.37**	**.28**	**.37**	**.40**	**.74**	**3124**	**.049**	**.920**

Note. Bold indicates the chosen model. Item numbers correspond to items in the GETS Inventory (see Appendix B). Briefly: Item 36 represents the use of simulation for teaching inference; Item 32 represents the requiring of explanations using written words on exam questions; Item 34 represents assessments completed in groups; Items 42 and 38 are the use of lectures and small-group activities to deliver content, respectively.
*Reverse coded

Model 2 was chosen based on lower AIC and acceptable measures of TLI and RMESA. Also, theory supported Model 2; the highest weighted item was regarding the use of activities; a reasonable indicator of student-centered beliefs (Kember, 1997). Coefficients from Model 2 were used as weights in calculating the Prior Beliefs measure. Scores were imputed for participants with 2 or fewer missing items (n=9), but participants with more than 2 missing items (n=27) were deemed unusable. Ten of these participants' responses were already unusable based on inability to calculate their score for Current Beliefs. As with the Current Beliefs measure, Prior Beliefs scores were re-centered and scaled to have mean of 0 and standard deviation of 1. The histogram of resulting scores is given in Figure 4. There is a clear mode between –1 and 0.

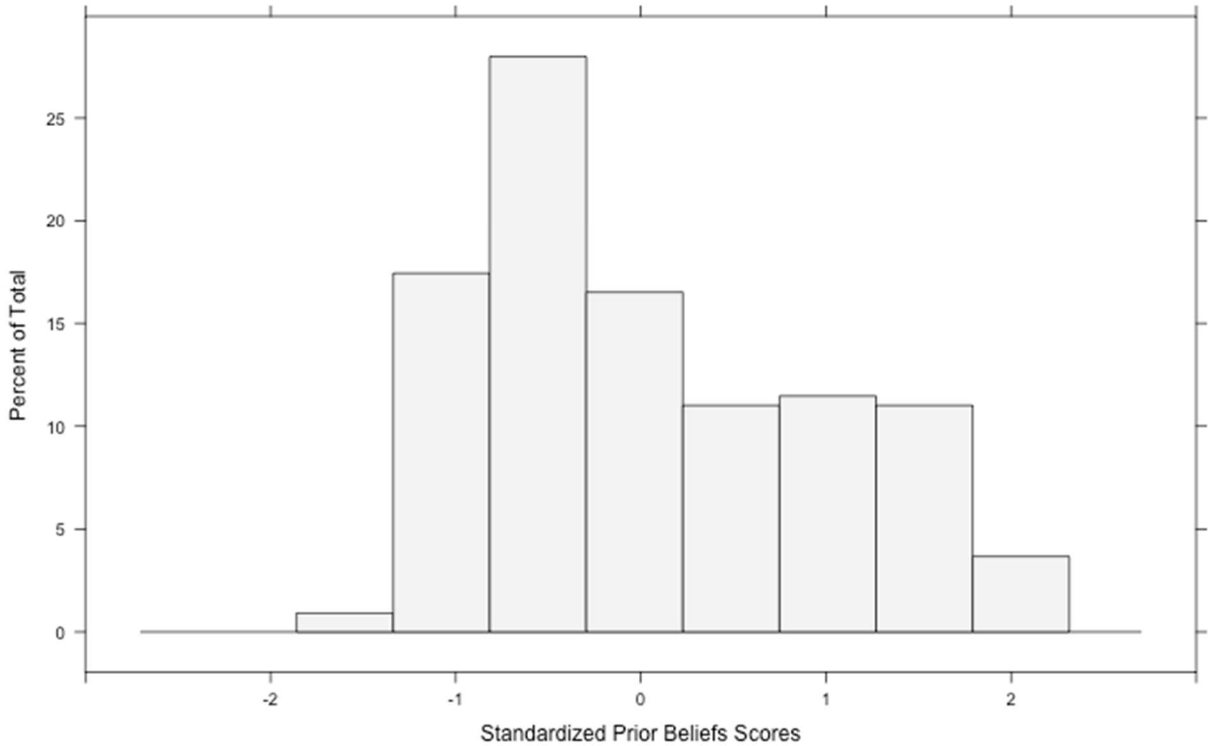


Figure 4. Distribution of scores for the measure of Prior Beliefs (*N*=218). Scores were weighted based on CFA coefficients, and standardized after imputation of appropriate missing values. These results should **not** be compared to those in Figure 3.

The scores for current and prior have been standardized within each respective set. Therefore it is **not** appropriate to compare the results of Prior Beliefs to those of Figure 3 to assess whether participants beliefs have changed over the course of their time in their current degree programs. A comparison of current and prior beliefs can be found in Subsection 4.1.1, Table 6 and Figure 2.

4.2.3 Measure of perceived engagement in the CoP. To develop a measure of participants' engagement in the CoP, several models were considered. Models used items regarding whether the participant has a shared office and if so the number of visits per week. Models were also considered that used items about whether participants felt comfortable approaching others (graduate students or faculty) in their program regarding teaching topics. All models also included two values indicating the extent to which graduate students participated in required and voluntary meetings.

Several steps were performed to calculate the values representing participation in required and voluntary meetings. First, responses to Items 14 and 18 (frequency of graduate students' attendance in voluntary and required meetings, respectively) were examined, and the last two response options were combined for each because there were so few participants who selected the most extreme high responses (*n*=2, and *n*= 11, respectively). Secondly the responses were re-coded as roughly the number of times per month that a meeting is held. For example, a response of "weekly or more" was coded as 4, "monthly" was coded as 1, and "Fewer than once per month" was coded as 0.5. For the voluntary meetings an additional option given was "more often than once per week," and this option was coded as 5. Thirdly, the coded participation values were multiplied by a transformation of the number of years that participants had participated in the meetings (Items 15 and 19, respectively). Specifically, a shifted log transform of the number of years the meetings had been attended was multiplied by the coded monthly frequency (x'=monthly frequency * log (years+1)). A log transform was used based on theory that assumed differences between small durations should carry more weight than differences between larger differences. The shift (+1) was so that 0 years of participation resulted in a score of 0, rather than an undefined score. The product of the monthly meetings times the transform of years produced the values that were used in all candidate models of Engagement as the number of graduate students' required and voluntary meetings.

Candidate models for the Engagement measure also included an indication of the number of people participants felt comfortable approaching regarding teaching topics (Items 21–22, Appendix B). Based on theory that engagement in CoPs can happen with as few as one other person, the results of these items were changed to be dichotomous before being included in candidate models. That is, for the faculty approachability measure, participants were scored as 1 if they indicated having at least one faculty member who is approachable to discuss teaching topics, and 0 otherwise. Similarly, for the student approachability measure, participants were scored as 1 if they indicated that at least one graduate student in their program is approachable to discuss teaching topics, and 0 otherwise.

Table 18 offers a summary of candidate models that were considered for the measure of perceived engagement in the CoP. As indicated by Table 18, all candidate models include scores for weekly meetings (required and voluntary) and weekly office visits. Candidate models differed in how they used the dichotomous measures of the extent to which faculty and graduate students are viewed as approachable for discussing teaching topics. Model 1 used a combined measure of faculty and graduate student approachability, which was the sum of the two dichotomous items described above. Model 2 and Model 3 used only faculty or only graduate students' approachability, respectively. Model 4 used both faculty and graduate students' approachability as separate topics. For the best performing model, Model 3, it was also considered whether office visits should be capped at 5 per week (i.e., grouping together the few 6 and 7 values with the 5's). This model (with the cap at 5) was the overall best performing model and was the chosen for the measure.

Table 18

Item CFA Coefficients and Fit Measures for Models of the Measure of Engagement in the CoP

	Topic								
				Approachability			Fit Measures		
Model	Required Meetings	Voluntary Meetings	Shared Office Visits	Faculty & Peers	Faculty	Peers	AICc	RMSEA	TLI
1	.55	.62	.51	.02	—	—	2728	.080	.853
2	.56	.61	.52	—	–0.02	—	2728	.091	.816
3	**.55**	**.62**	**.51**	**—**	**—**	**.05**	**2727**	**.048**	**.946**
4	.55	.62	.51	—	–0.01	.05	3426	.159	.326

Note. Based on N=218 participants whose responses were useable in the final model. Bold indicates final chosen model. Topics are based on items given in Appendix B.

Coefficients from Model 3 were used to calculate a weighted score for the measure of participants' perceived engagement in the CoP. The results were scaled to have a mean 0 and standard deviation 1. Missing data were not an issue and no imputations were necessary. Scores for engagement are given in Figure 5. The distribution is unimodal, with some slight skew to the right. The right skew is not surprising, as it is possible for some participants to be extremely active and engaged in CoPs.

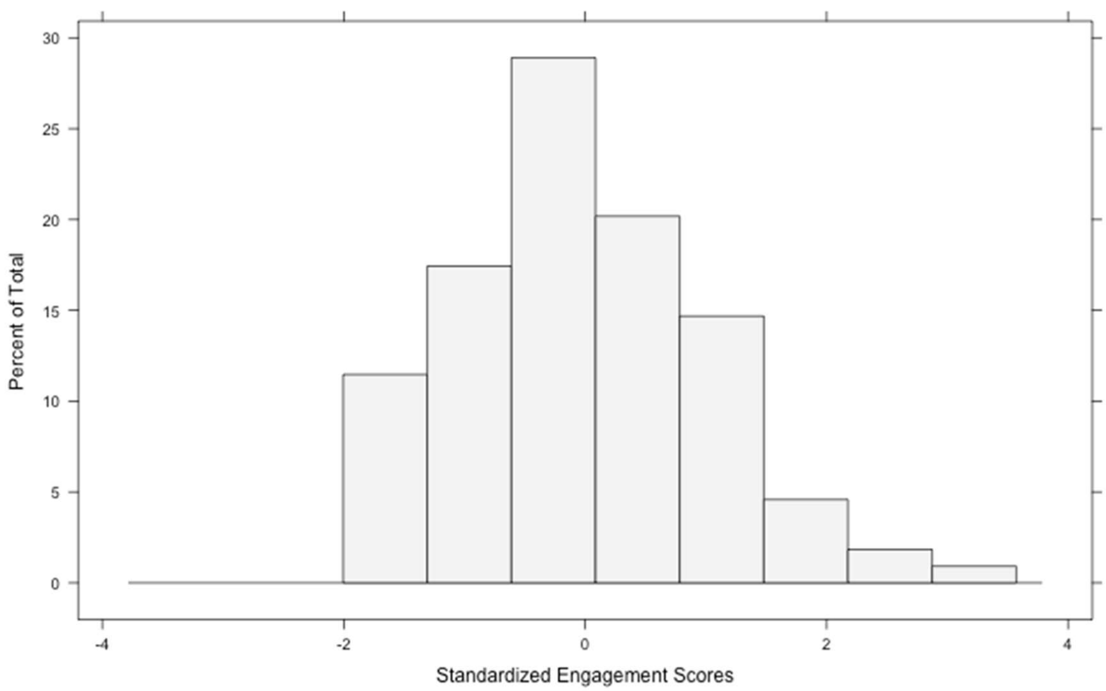


Figure 5. Distribution of scores for measure of Engagement in CoP. Based on *N*=218 participants included in the final model.

4.2.4 Measure of norms of interaction in the CoP. To develop a measure of the perceived norms of interaction in participants' CoPs, Items 47 – 56 were considered for use in candidate models. These dichotomous items asked about participants' interactions with the two people they identified as most influential upon their beliefs about teaching statistics. As indicated by the responses in Appendix B, there was not much variation among responses regarding the CoP Norms of interaction. That is, for most participants norms appeared to be very positive for all items. This lack of variation proved problematic when defining a measure. No models were able to produce adequate model fit indices.

The four models that were considered and their fit measures are given in Table 19. Model 1 uses all of Items 45—55. Model 2 does not use Items 49 and 55 because both items negatively correlated with the rest. Also, this item was not rooted in the literature and had less theoretical justification as a measure of faculty contributions to graduate students' teaching experiences. Models 3 and 4 used items based on the department members chosen as first and second most influential, respectively. The polychoric correlation matrix that was computed using the items regarding whether participants admired the influential people (Items 49 and 55) was not positive definite. Once the problematic admire item was removed the polychoric correlation matrices were positive definite.

Table 19

Coefficient Estimates and Fit Measures for Models of Norms of Interaction Measure.

	Item Number										Fit Measures		
Model	47	48	49	50	51	52	53	54	55	56	AICc	RMSEA	TLI
1	.67	.74	.68	.65	.42	.69	.84	.80	.51	.44	5981	.773	–.09
2	.63	.69	.57	.58	—	.62	.75	.69	.45	—	4683	.464	.22
3	**.73**	**.92**	**.55**	**.89**	—	—	—	—	—	—	**2247**	**.159**	**.93**
4	—	—	—	—	—	.48	–.17	.44	.90	—	2223	.528	.39

Note. Based on *N*=218 participants whose responses were useable in the final model. Bold indicates final chosen model. See Appendix B for items. Briefly, Items 47 & 52 are regarding Comfort engaging in discussion; Items 48 & 53 represent acknowledgement of point of view; Items 49 & 54 are regarding respect; Items 50 & 55 are regarding care for the quality of teaching; and Items 51 and 56 are regarding admiration of the influential person. The item with the earlier number is for the department member chosen as most influential, while the second item regarding each topic is regarding the department member chosen as the second most influential.

The fit measures for Model 3 do not indicate adequate fit by typical standards for RMSEA (e.g., Hu & Bentler, 1999; RMSEA<.06), however Model 3 was chosen because it has the best AICc and TLI. Also, Model 3 fits most consistently with theory; the admire item was not expected to be as consistent with the measure. The coefficients for Model 3 were used to calculate score for the measure of engagement. Scores were not imputed for those with missing data (n=2) because the measure did not appear helpful in later analyses.

The distribution of scores for the measure of norms of interaction is given in Figure 6. There is a clear mode on the right corresponding to mostly positive interactions between participants and the people chose as most influential on their teaching beliefs. Meanwhile, a few participants indicated relatively very negative experiences with the people who they perceive to be most influential.

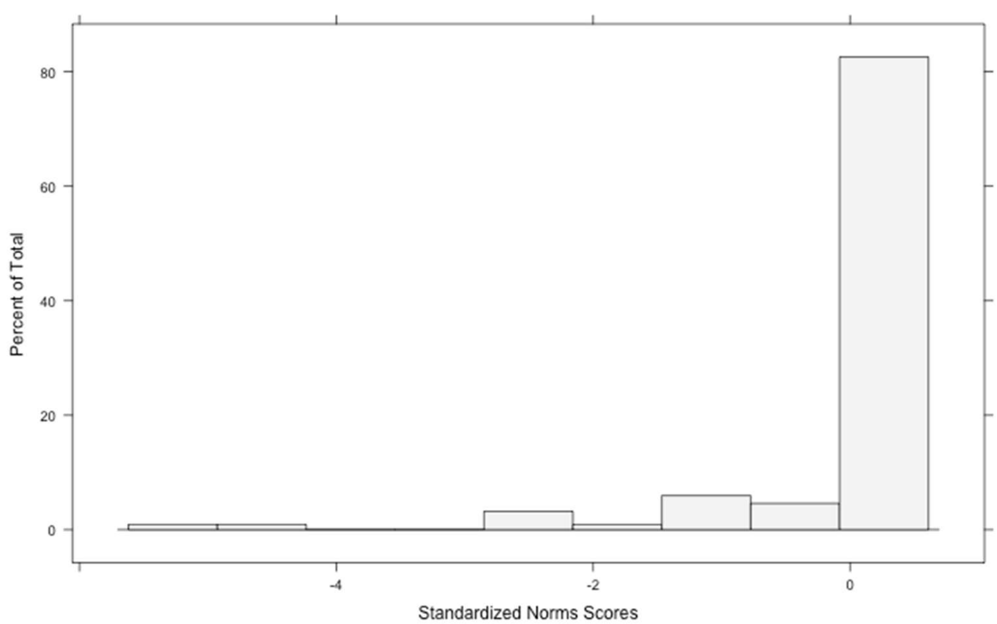


Figure 6. Distribution of Scores for measure of Norms of Interaction in CoP. (N=216) .

4.2.5 Measure of perceived CoP orientation toward student-centered teaching. A measure of graduate students' CoPs' orientation toward student-centered teaching was created using the six dichotomous items about the teaching practices of the department member that participants perceived to be most influential (Items 57–62, Appendix B). Using these items, a polychoric correlation matrix was examined to determine items that were candidates for removal from the model. Items that did not produce much variation among participants were also considered for removal. For example, once the data were collected it was discovered that the simulation question may be too strong ("primary use") to be useful in measuring variation for the current set of participants. The same argument rendered the item about individual grades eligible for removal from the model. Table 20 gives the three models that were considered, and the fit measures used to evaluate them.

Table 20

Coefficient Estimates and Fit Measures for Models of CoP Orientation Measure

	Item CFA Score Coefficients Item Number						Fit Measures		
Model	57	59	61	58	62	60	AICc	RMSEA	TLI
1	.81	.70	.61	.46	.69	.28	3796	.244	.538
2	.86	.66	.59	.45	.67	—	3119	.187	.770
3	**.81**	**.70**	**.67**	**.38**	—	—	**2538**	**.126**	**.897**

Note. *N*=217. Briefly, Item 57 = use of lectures, Item 59 = use of activities, Item 61 = use of a variety of means and media for students to communicate their ideas. Item 58 = frequent practicing of procedures using formulae, Item 62 = use of individual assessments, and Item 60 = use of simulation methods to teach statistical inference.
*Reverse coded

Model 3 was chosen because it had the best AICc, the TLI indicated near adequate fit, and the RMSEA was the best of the candidate models. Also, the coefficients of Model 3 align with theory; the heaviest weights are for use of lectures and activities, which are common indicators of teacher-centered and student-centered teaching (Kember, 1997). Based on Model 3, participants' orientation scores were based on reported community members' use of lectures (reverse coded), activities, opportunities for students to communicate ideas, and frequent practice using formulas (reverse coded).

The coefficients for Model 3 were used to calculate a score for the measure of orientation. Six participants had missing data, four of which were already deemed ineligible from non-response for the beliefs construct. Of the remaining two, scores were imputed for the participant that was missing one out of the four items. The other participant was deemed ineligible, missing all four items. Scores were recentered and rescaled to have mean 0 and standard deviation of 1. The distribution of scores for CoP orientation, which is unimodal and somewhat bell-shaped, is given in Figure 7.

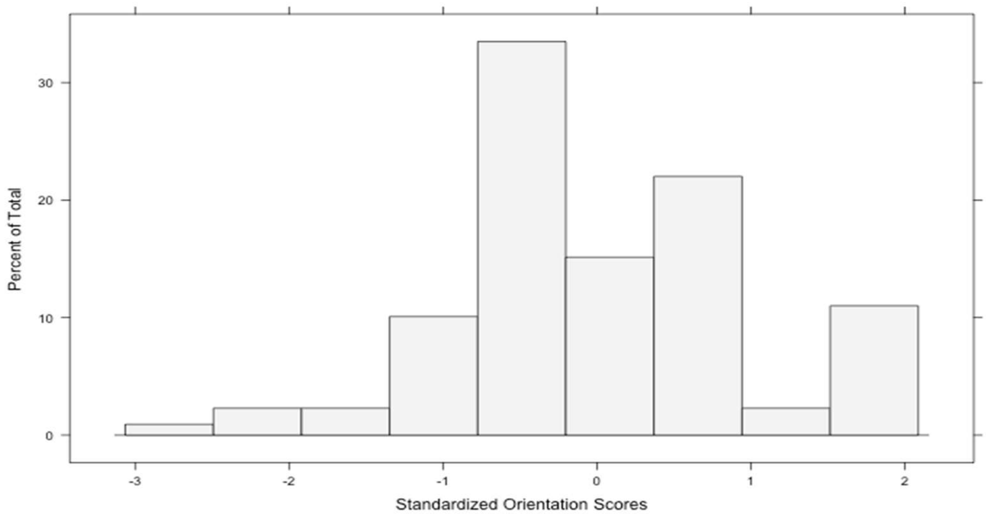


Figure 7. Distribution of scores for measure of perceived Orientation of CoP. (*N*=217)

The perceived orientation of the most influential person does not appear to be associated with the role of that person in the department. Figure 8 gives the orientation scores conditional on whether the most influential person is a faculty member or not. The distributions to not appear to be very different.

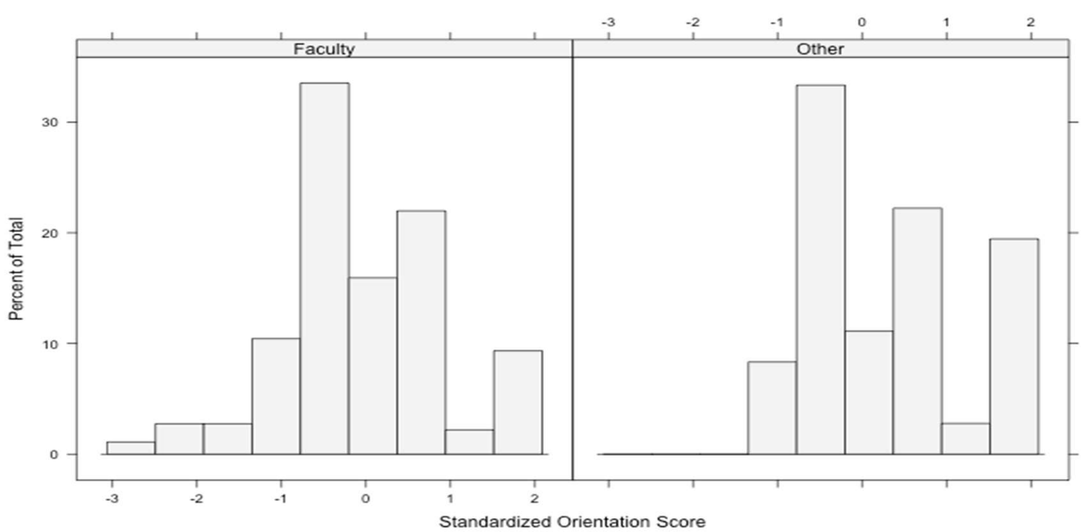


Figure 8. Orientation scores conditioned upon role of most influential person (faculty or not faculty). *N*=217.

4.2.6 Measure of perceived faculty presence in CoP. Candidate models for the measure of perceived faculty presence are based on several items from the survey instrument. All items used for this measure needed to be prepared before being suitable for inclusion in the model. For example, to temper for outliers, the number of faculty members that participants indicated they would feel comfortable approaching with teaching questions (Item 22) was re-coded as 0,1,2, or 3 (for 3 or more). Candidate models also included a measure of the number of times faculty provided feedback after observing graduate students performing their teaching-related duties (Item 26). Also to temper outliers, the number of observations was recoded as 0, 1, 2, or 3 (for 3 or more). The number of faculty members chosen as one of the two primary influencers for Items 45 and 46 was summed and included as 0,1, or 2 faculty influencers.

All models included an indication of the extent to which faculty participated in mandatory and voluntary meetings. Several steps were performed to calculate these indicators, and the first two steps match those of the Engagement measure described earlier. As with the Engagement measure, responses to Items 14 and 18 (frequency of graduate students' attendance in voluntary and required meetings, respectively) were examined, and the last two response options were combined for each because there were few participants who selected them (n=2, and n= 11, respectively). Also, the responses were re-coded as roughly the number of times per month that a meeting is held (Items 14 and 18, respectively). Thirdly (now different from the Engagement measure) the frequency of meeting codes were multiplied by the percent of meetings attended by faculty (Items 16 and 20, respectively). The product was used in all candidate models for the number of graduate students' required and voluntary meetings attended by faculty, per month.

The last set of items that required preparation before being included in candidate models were the items designed to measure behaviors about faculty support for graduate students. The three behaviors that were used, which were gathered from the literature regarding faculty support for graduate students, are given in Items 23–25 (Appendix B). Some models included the three items separately, and other models used the sum of items for which the response was positive. When the models that used the sum outperformed the models that treated the behaviors individually, only the sum models were considered in candidate models and reported.

Table 21 gives the three candidate models, coefficient estimates, and fit measures. Model 3 was chosen for the faculty participation measure. Although Model 1 had better RMSEA and Model 2 had higher TLI, Model 3 had the lowest AICc and other fit measures (TLI, RMSEA) were adequate. Also coefficients of Model 3 were all in the appropriate directions according to theory. The coefficients from Model 3 were used as weights to calculate a faculty participation score. There were no missing data for this measure.

Table 21

Coefficient Estimates and Fit Measures for Models of Faculty Participation Measure

	Topic						Fit Measures		
Model	Required Meetings	Voluntary Meetings	Observation & Feedback	Behaviors [a]	Influential [b]	Approachable [c]	AIC	RMSEA	TLI
1	.50	.41	.49	.56	.05	.04	4113	.023	.973
2	.50	.41	.49	.60	.05	—	3415	.028	.973
3	**.49**	**.41**	**.49**	**.61**	—	—	**2717**	**.037**	**.972**

Note. *N*=218. All topics are based on one or more items that have been transformed or recoded to taper the effects of outliers.
[a] Sum of the three dichotomous faculty behavior items (Items 23-25).
[b] Number of faculty members chosen as one of the two primary influencers (Items 45 & 46).
[c] Number of faculty members that participants indicated they would feel comfortable approaching with teaching questions (Item 22), re-coded as 0,1,2, or 3 (for 3 or more).

Figure 9 gives the distribution of the standardized Faculty Presence scores. The distribution is skewed to the right. The right skew is not surprising, as there is a lower limit to the possible values of how involved faculty may be in the CoPs.

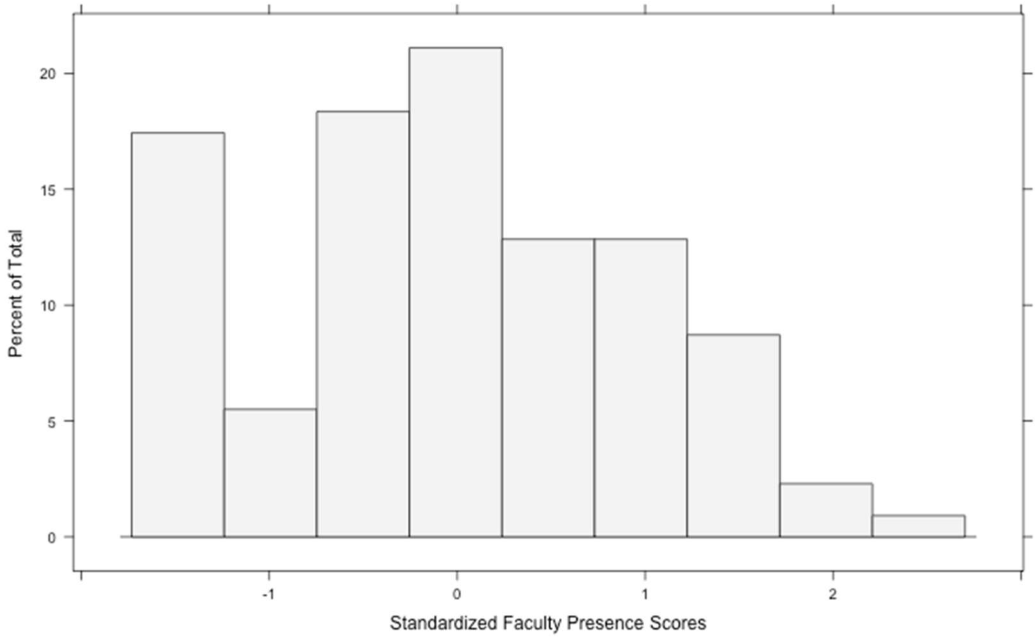


Figure 9. Distribution of scores for measure of perceived Faculty Presence in CoP. (*N*=218)

4.3 Path Analysis

Once measures were calculated for the six core constructs of the study, path analysis was conducted to investigate the relationships between the constructs. All models were calculated using a correlation matrix of pairwise complete observations (see Appendix L). Path analysis was conducted using the lavaan package in R (Rosseel, 2012).

Paths were removed when the coefficients were the opposite sign from what theory would suggest, as suggested by Keith (2006, p. 273) for exploratory research.

Also, paths were removed for the construct that did not have enough variation from participant responses to warrant proper investigation. For the remaining models, the fit measures used to compare and evaluate models were the AIC, TLI, and RMSEA. For the larger models, there were *N*=214 cases. As variables were eliminated, there were fewer respondents with missing data, so the useable sample size grew to *N*=218.

Table 22 offers the path coefficient estimates and fit measures for each of the models that were tested. The first candidate model was the full theoretical model given in Appendix A. Because of the lack of variation in the Norms measure and the negative paths the measure produced, the three paths associated with Norms of Interaction were the first to be removed (Model 2). For Model 3 and Model 4 two more paths were removed because the coefficients did not appear to be very large. Finally, two more paths were removed because the engagement path was not significant. Removal of this path left the path leading to engagement irrelevant for this study. The final model, and the model with the best-fit measures, was Model 5, which included only the paths from the Prior Beliefs and the CoP Orientation constructs to the Current Beliefs construct.

Table 22

Path Analysis Coefficients and Fit Measures

	Path									Fit Measures		
Model	d	e	c	g	i	a	b	f	h	AIC	RMSEA	TLI
1	.06	.48	.15	.68	.01	–.01	.04	.18	–.12	3497	.086	.89
2	.07	.48	.15	.67	.01	.00	—	—	—	2903	.000	1.030
3	.07	.48	.15	.66	.01	—	—	—	—	2900	.000	1.035
4	.07	.48	.15	.66	—	—	—	—	—	2895	.000	1.034
5	**.07**	**.49**	—	—	—	—	—	—	—	**1791**	**.000**	**1.000**

Note. Fit measures are based on correlations from pairwise complete observations. For Model *N*=214 complete cases, for all other models *N*=217. See model in Appendix A for relationships and path identification.

In these preliminary analyses it was found that the best fit measures corresponded with a model for which path analysis is not necessasry. Model 5 has two (unrelated) predictors. All other models had AIC values that were greater by more than 20, rendering them implausible (Burnham, Anderson, & Huyvaert, 2011). In light of the lack of structure, path analysis was deemed unnecessary, and subsequent analyses used cross-validation techniques to evaluate ordinary least squares regression models.

4.3.1 Preparation for cross-validation and model selection. Before discussing the cross-validation results, briefly here we offer the theoretical rationale for the candidate base models. These models were called candidate *base models* because they contained no variables other than the main constructs of the study (e.g., Current Beliefs, Prior Beliefs, Engagement). No characteristics variables (e.g., year in program, interest in teaching) were included in base models.

As recommended by Burnham & Anderson (2004), efforts were made to reduce the number of candidate models for cross-validation and model selection. Interactions were limited to two-variables at a time. Whenever interactions were included, so also were the corresponding main effects. For all models, a main effect of Prior Beliefs was included as a covariate. None of the models use the Norms variable because there was not enough variation in the measure to justify its use.

Based on theory, most variables were not considered for main effects alone. For example, theory would suggest that the Faculty Presence effect on Current Beliefs would be mediated by Orientation; a strong faculty presence could be very student-centered or teacher-centered. Therefore, models that include Faculty Presence variable also include the interaction of Faculty Presence and Orientation. Similar arguments were made for the

Engagement variable; a graduate student could be very engaged in a CoP that is either student-centered or teacher centered. Therefore Engagement was always included with an interaction with Orientation. Also, reciprocally, the Orientation variable was included only if interacting with Engagement or Faculty Presence. Theoretically, the Orientation would not have much impact if the participant does not engage in the CoP or with faculty.

After taking these theoretical arguments into consideration, there remained four candidate base models. The models include the four possible combinations of including the Orientation-Engagement interaction, the Orientation-Faculty Presence interaction, neither, or both interactions. Variables were centered to avoid colinearity, and the VIF was checked for values less than 10 when more than one predictor was used. All models included Prior Beliefs as a covariate. The four candidate base models are given below in order of increasing complexity.

$$\widehat{\text{Beliefs}}_1 = \beta_0 + \beta_1 \text{Prior Beliefs}$$

$$\widehat{\text{Beliefs}}_2 = \beta_0 + \beta_1 \text{Prior Beliefs} + \beta_2 \text{Orientation} + \beta_3 \text{Engagement} + \beta_4 \text{Orientation} * \text{Engagement}$$

$$\widehat{\text{Beliefs}}_3 = \beta_0 + \beta_1 \text{Prior Beliefs} + \beta_2 \text{Orientation} + \beta_3 \text{Faculty Presence} + \beta_4 \text{Orientation} * \text{Faculty Presence}$$

$$\widehat{\text{Beliefs}}_4 = \beta_0 + \beta_1 \text{Prior Beliefs} + \beta_2 \text{Orientation} + \beta_3 \text{Engagement} + \beta_4 \text{Faculty Presence} + \beta_5 \text{Orientation} * \text{Engagement} + \beta_6 \text{Orientation} * \text{Faculty Presence}$$

Model assumptions were examined, including checks for normally distributed residuals (e.g., examination of histograms and QQ-plots of residuals) and checks for homoscedasticity of residuals (e.g., examination of plots of residuals against fitted values). Candidate models were also examined for non-linear relationships using visual examination of added variable plots. When no plots revealed curves that would suggest transformations were necessary, the process continued. Plots of pairwise relationships between the five variables did not reveal any signs of non-linear relationships, so no transformations or higher-order predictors were used. Also, OLS regression was used to check that the coefficients matched theory. For all candidate base models the assumptions appeared to be reasonably met. The only noteworthy issue may be the ceiling effects that limited the variation in residuals for participants who earned the maximum possible score for student-centered beliefs.

4.4 Cross-Validation and Model Selection

Cross-validation techniques with ordinary least squares linear regression were used to explore models of graduate students' beliefs about teaching statistics. Cross-validation techniques were used to protect against overfitting (Breiman, 2001). The model selection process occurred in two stages. First, candidate base models were explored and a base model was selected. Secondly, characteristics supported by theory and prior research (e.g., year in program, international student status) were added to the base models and explored using cross-validation.

The primary fit measure used for model selection was the corrected Akaike information criterion (AICc). The AICc (corrected) was used instead of the AIC (not corrected) because the largest candidate model had sample-size-to-parameter ratio that did not exceed 40 (Burnham & Anderson, 2004). Also reported was the Bayesian information criterion (BIC). However AICc was considered more appropriate than BIC for this study because beliefs are complicated constructs (Pajares, 1992) and are likely described by many tapering effects, rather than a few large effects (Burnham & Anderson, 2004; Vrieze, 2012; Weakliem, 1999). As suggested by Burnham, Anderson, and Huyvaert (2011), models with AICc within 7 of the minimum were retained.

In addition to the AICc and BIC, the mean squared error (MSE) from the cross-validation results was reported. To be specific, the MSE, ${\sigma_\varepsilon}^2$, used the squared prediction errors $(\hat{y}_i - y_i)$ of the training model applied to the validation set. Namely, for each fold, within each iteration, the MSE was calculated as

$${\sigma_\varepsilon}^2 = \sum_{i=1}^{n_k} \frac{(\hat{y}_i - y_i)^2}{n_k},$$

where n_k is the number of cases in the test set when the data are divided into *k* folds. The final reported average MSE, $\overline{{\sigma_\varepsilon}^2}$, is the average across the *k* folds for the *j* iterations:

$$\overline{{\sigma_\varepsilon}^2} = \frac{1}{jk} \sum_{v=1}^{j} \sum_{w=1}^{k} ({\sigma_\varepsilon}^2)_{vw}.$$

According to Zhang and Yang (2015) and contrary to some recommendations, cross-validation using only 10-fold resplittings is not always appropriate. Therefore analysis of the base models was conducted using four different splitting ratios, namely half-half (2-fold), 5-fold, 10-fold, and 20-fold. For each model and splitting ratio, Monte-Carlo cross-validation was conducted for *j*=1000 resplittings. That is, the data were re-allocated into the training and test sets 1000 times.

To conduct the cross-validation a series of functions were written in R (available in Appendix M). Inputs of the function were: the model to be tested, the data set, and the number of folds (*k*) used, and the number of resplittings (*j*). Before completing calculations, the function removed all cases from the data that had missing values for the specified model, and randomly assigned the data into groups corresponding to the splitting ratio. For example, for a 5-fold ratio, data were divided into $k = 5$ groups. Fit measures were computed using each of the *k* sets as test-sets. Then the process was re-randomized and repeated *j* times, resulting in a total of $j*k$ values for each fit measure. The average and standard deviation were reported for each fit measure.

4.4.1 Cross-validation and model selection: base model. The results of the cross-validation for the four models for each of the splitting ratios ($k = 2, 5, 10, 20$) are given in Table 23. The average and standard deviation of the three model fit measures are reported, and, in the case of AICc and BIC, the difference from the best-performing model is also reported. The smallest AICc and BIC indicate better fit, so negative values with large magnitude indicate better AICc and BIC.

Table 23

Cross-Validation Results of 1000 Resplittings for Candidate Base Models.

	MSE		AICc			BIC		
Model	Mean	SD	Mean	SD	$\Delta_{\bar{X}_i-\bar{X}_{min}}$	Mean	SD	$\Delta_{\bar{X}_i-\bar{X}_{min}}$
				2-fold				
1	**.77**	**.07**	**–24.7**	**9.4**	**0**	**–19.4**	**9.4**	**0**
2	.78	.07	–17.2	9.7	7.5	–4.4	9.7	15.1
3	.78	.07	–16.7	9.2	8.0	–3.9	9.2	15.5
4	.80	.07	–10.2	9.8	14.6	7.5	9.8	27.0
				5-fold				
1	**.79**	**.05**	**–36.1**	**10.5**	**0**	**–29.9**	**10.5**	**0**
2	.84	.09	–19.7	16.6	16.4	–4.3	16.6	25.6
3	.84	.08	–20.3	15.1	15.8	–4.9	15.1	25.0
4	.90	.11	–4.9	20.4	31.2	16.5	20.4	46.4
				10-fold				
1	**.83**	**.08**	**–32.7**	**18.0**	**0**	**–26.2**	**18.0**	**0**
2	1.02	.27	8.4	40.9	41.1	24.4	40.9	50.1
3	1.00	.24	6.0	38.3	38.7	22.1	38.3	48.3
4	1.23	.46	46.0	55.8	78.7	68.3	55.8	94.5
				20-fold[a]				
1	**.92**	**.23**	**–16.0**	**37.1**	**0**	**–9.4**	**37.1**	**0**
2	2.01	4.59	102.9	113.9	118.9	119.3	113.9	128.7
3	1.95	5.49	98.3	110.3	114.3	114.7	110.3	124.1
4	10.00	317.7	238.2	175.7	254.2	261.0	175.7	270.4

Note. N=217. Bold indicates the best performing model.

[a] Some 20-fold results were rank-deficient (i.e. not enough cases in the test set to appropriately assess fit measures) and should be treated with caution.

For all splitting ratios, the model with the best average validation error, AICc, and BIC was Model 1. The standard deviation of values for Model 1 is relatively small relative to the standard deviations for the other models, suggesting that for many resplittings of the data, Model 1was the best performing model. Model 1 uses just prior beliefs as a covariate and none of the other core measures (e.g., Orientation, Faculty Presence, Engagement).

Using the cutoff of AICc within 7 of the minimum AIC (as recommended by Burnahm et al., 2011), no other candidate models were retained and used for the second stage of model selection. Model 1 was used as the base model for the next section, where the model was explored for invariance under other graduate student characteristics.

4.4.2 Cross-validation and model selection: models with characteristic variables. After the base model was chosen, cross-validation methods were used to explore main effects of graduate student characteristics over and above the chosen base model. Only a half-half (two-fold) splitting ratio was used because some of the characteristics were not evenly split across participants (e.g., prior teaching experience in primary or secondary education) and did not have enough participants in the smaller groups to support higher splitting ratios. To limit the number of models considered (Burnham & Anderson 2004), only main effects were used for characteristics variables, and the variables were included only one at a time. In total, ten characteristic variables were investigated for inclusion with Model 1.

Some of the characteristics were selected to be explored based on prior research on GTAs (e.g., international student status (dichotomous), prior experience teaching primary or secondary levels, prior college-level teaching experience). Others characteristics were guided by scholarship, theory, and seemed logical to investigate (e.g., department type (dichotomous; biostatistics or statistics), interest in teaching as part of career (dichotomous), year in the program (0–1,2, or greater than three), plans to earn a Ph.D. at current institution (yes, undecided, no), experience as a research assistant (dichotomous), experience as a graduate instructor (primary teacher for a course; dichotomous), and experience with any teaching-related responsibilities (dichotomous; none, or any of grading papers, holding office hours, assistant to a primary instructor, facilitating lab or discussion sections, or serving as a primary instructor).

As with the base models, for each model and splitting ratio, cross-validation was conducted for j=1000 resplittings. The results are presented in Table 24. The average and standard deviation of the three model fit measures are reported, and, in the case of AICc and BIC, the difference from the best-performing model was also calculated and reported.

Table 24.

Two-fold Cross-Validation Results for j=1000 resplittings of Models With Characteristic Variables

Characteristic	MSE		AICc			BIC		
	Mean	SD	Mean	SD	$\Delta_{\bar{X}_i - \bar{X}_{min}}$	Mean	SD	$\Delta_{\bar{X}_i - \bar{X}_{min}}$
Department type	.78	.07	–21.3	9.6	3.4	–13.5	9.6	5.9
PhD in current dep	**.76**	**.07**	**–22.4**	**9.8**	**2.3**	**–12.1**	**9.8**	**7.4**
Interest in teaching	.78	.07	–21.0	9.6	3.7	–13.2	9.6	6.3
International status	.77	.07	–22.8	9.7	1.9	–15.0	9.7	4.4
Year in program*	.77	.07	–20.3	9.9	3.9	–9.9	9.9	6.4
Prior experience…								
Teaching K–12	.78	.07	–21.3	9.4	3.4	–13.5	9.4	6.0
Teaching coll-level	.78	.07	–20.8	9.6	3.9	–13.0	9.6	6.4
Experience in current program…								
Teaching/assisting	.78	.07	–21.8	9.7	2.9	–14.0	9.7	5.5
Instructor of record	.78	.07	–21.6	9.7	3.1	–13.7	9.7	5.7
Research assistant	.77	.07	–23.1	9.8	1.6	–15.2	9.8	4.2
***None* (Base Model)**	**.77**	**.07**	**–24.7**	**9.4**	**0**	**–19.4**	**9.4**	**0**

Note. N=218. Bold indicates the model with the best average MSE (phd model), or best AICc and BIC (base model). All models (including the base model) use prior beliefs as a covariate.
*Year in program treated as a factor with 3 levels: 0–1, 2, or greater than 3.

As recommended by Burnham et al. (2011), models that produced AICc within 7 of the lowest were retained as plausible. Models were also checked for coefficients that matched theory. Based on these criteria, all the candidate variables appear in plausible models. Models that included main effects for year in program, experience hired as a research assistant, intent to earn a doctoral degree at current institution (as opposed to a master's degree only), international student status, department type (statistics or biostatistics), prior teaching experience, future plans to teach as a career, and experience with teaching responsibilities, or none of the characteristics above, were all plausible.

4.4.3 Final (base) model. Although many models that include characteristic variables were considered plausible, the primary base model that arose from cross-validation results is a simple linear model predicting current beliefs from prior beliefs. Figure 10 offers a plot of the data and the line of best fit. The plot summarizes the primary relationship that was found in this study: variation in current beliefs can be explained, in part, by prior beliefs about teaching statistics. The data do not suggest any strong curvilinear relationships, although there is a clear ceiling effect that may be restricting what may have been more natural variation on the top right hand side.

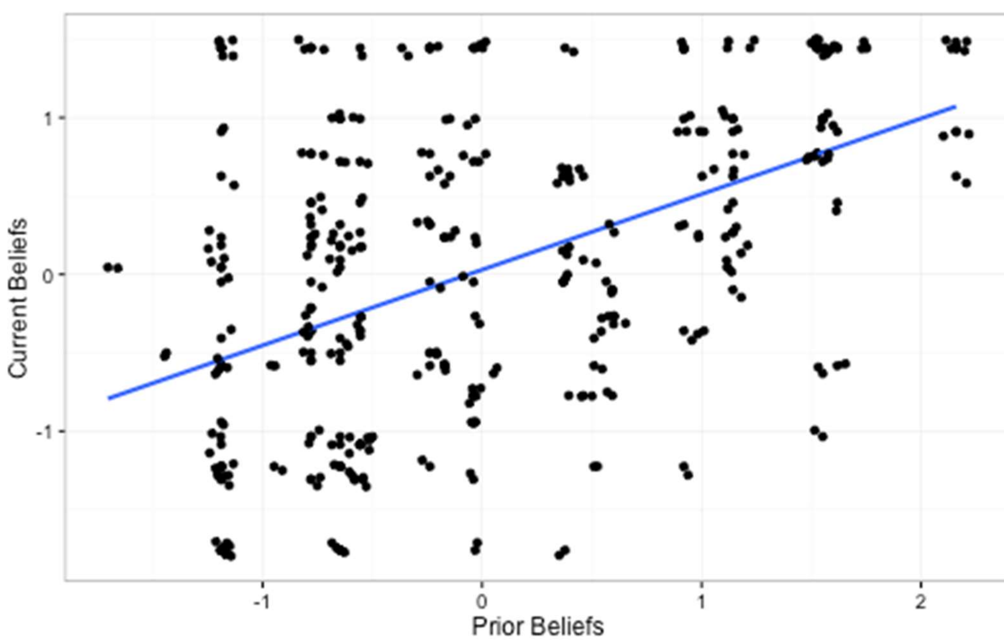


Figure 10. Plot of (jittered) Current Beliefs vs. (jittered) Prior Beliefs, with Model 1 overlaid. Based on *N*=218 cases. A ceiling effect appears to be at play, particularly for those with high scores for prior beliefs.

4.5 Chapter Summary

This chapter described the survey results and data analyses. The sample included statistics and biostatistics graduate students from 37 institutions across the United States. Confirmatory factor analysis results were used to create measures for the six major core constructs of the study. The measures were used to evaluate models of relationships between the core constructs of the study. The data did not support the proposed theoretical model of relationships between the six core constructs. Using cross-validation techniques, a final "base" model was chosen and explored for invariance across other graduate student characteristics. In the final chosen base model Prior Beliefs was the only core construct used to predict Current Beliefs. However, many characteristics of graduate students appear to have potential for explaining additional variation in statistics graduate students' beliefs about teaching statistics. Chapter 5 offers a discussion of these results.

Chapter 5

Discussion

This study was conducted to explore the relationship between graduate students' experiences in CoPs and their beliefs about teaching statistics. A survey was created and administered to statistics graduate students across the United States. The results of the survey were used to define measures for the six main constructs of the study. Four of the constructs represent aspects of GTAs' experiences of CoPs, and two constructs relate to GTAs' teaching beliefs. Theoretical models were explored to examine relationships between the constructs.

Due to the fact that the sample was not randomly selected from the population of GTAs in United States statistics departments, the responses are not generalizable to all statistics graduate students in the nation. Results are discussed within this study's particular context (Guba, 1981; Shenton, 2004). For example, because the invitation was made available via faculty it is plausible that the sample is biased toward of graduate students who have positive relationships with faculty members. Also, the sample may be representative of students who are interested education because the survey was advertised to focus on topics in this area (see Appendices F–J). Most of the participants are within their first two years of entering their current degree programs, so the sample may not adequately represent graduate students who have made further progress in their degree programs.

However, there are aspects of the sample that suggest it is representative within the contexts given above. Institutions from all major geographic regions of the United States are represented. Public and private institutions are represented, as well as a variety of different sizes of institutions (Appendix K). Also, participants represent a variety of demographic backgrounds (Sections 6, Appendix B). For example, there is a fairly even split of gender, with a few students who indicated they do not identify as male or female. Also, about a quarter of participants indicated that they are international students.

This chapter offers a discussion of the study's contributions to research about statistics graduate students' beliefs about teaching statistics and their participation in CoPs. Also included is a discussion of the GETS Inventory and its contributions in areas of measuring teaching beliefs and measuring participation in CoPs. The chapter finishes with limitations of the study and implications for future research.

5.1 Statistics Graduate Students' Beliefs about Teaching Statistics

The results of this study suggest that participants' beliefs about teaching statistics tend to fall short of student-centered teaching as defined by Kember (1997) and recommended by scientific organizations (e.g., AAAS, 1989; ASA, 2016). For example, when asked about teaching beliefs for a hypothetical class of about 35 students, most participants indicated that less than a quarter of class time should be used for students to work together in small groups. Participants also indicated that less than half of assessment questions should require explanations using words. On average, participants indicated that the instructor should lecture for about 60% of class time. These teacher-centered results are consistent with previous research that suggests that the majority of statistics graduate student are not aware of professionally endorsed guidelines for teaching statistics and have not learned about research on how students learn statistics (e.g., Justice et al., in press).

It appears that many participants' teaching beliefs have become more student-centered since entering their current degree programs. More than half of participants indicated less use of lecture and greater use of small-group activities than what they believed to be appropriate when they entered their current programs. Also the majority indicated that homework problems should involve less practicing procedures using formulae. It is unclear why participants' beliefs have become more student-centered. Perhaps their departments have modeled student-centered teaching practices (e.g., Rumsey 1998). Another possible reason is that participants have engaged in courses that encourage them to use student-centered teaching practices (e.g., Garfield & Everson, 2009).

Unfortunately, the changes in the student-centered direction are often small. As indicated by Figure 2 in Chapter 4, the typical percentage difference was between 0 and 20%. This result is consistent with previous research that suggests beliefs about teaching are often strongly held and resistant to change (e.g., Fang, 1996; Kane et al., 2002; Pajares, 1992).

This study did not provide evidence to suggest that students' year in their program was associated with the extent to which their teaching beliefs have changed. For each topic about teaching beliefs included in the survey, there was not a significant relationship between participants' year in program and the magnitude of their change in beliefs. This result is surprising when compared to previous research (e.g., Wyse, 2010) that suggests changes in beliefs need time (e.g., more than one semester) to be detectable. One may think that students who have had more years in the program may have larger changes in beliefs. However for this sample there was not evidence of an association between time in program and magnitude of change.

5.2 Statistics Graduate Student Participation in CoPs

This study gathered data regarding the nature and extent of graduate students' participation in CoPs. Overall it appears that most participants have access to some form of a CoP. Three-fourths of participants who have been assigned teaching responsibilities indicated they have been required to attend meetings regarding teaching topics. About three-fourths of all participants indicated they engage in voluntary discussions regarding teaching topics as well. Nearly all participants feel that there are faculty or graduate students in their departments that they are comfortable approaching to discuss teaching topics.

Participants do not appear to have participated in CoPs for very long. Over half of the participants who have been required to attend meetings indicated that the meetings have lasted for one year or less. About half of respondents who participate in voluntary meetings have done so for no more than one year. For the one quarter of participants who have been in their degree programs for one year or less, the short duration of CoP participation is to be expected. For the remaining three-fourths of respondents, it is uncertain why their participation in CoPs seems to be fairly short.

When compared to previous research about faculty presence, the results of this study are somewhat surprising. In this sample, 83 percent of participants indicated that a faculty member was the most influential person in their department with respect to their teaching beliefs. Nearly half of participants chose faculty members as their second most influential department member as well. This result complements research that suggests that GTAs socialize one another (e.g., Darling, 1989). Although GTAs may find each other to be most helpful and most available (e.g., Myers, 1996), the results of this study suggest graduate students perceive faculty—not other GTAs—to be most *influential* upon their teaching beliefs.

Perhaps less surprising is the evidence to suggest that participants typically feel comfortable, respected, free to disagree, and cared about by the people they feel are most influential regarding their teaching. At least 90% of participants gave positive responses on the five items regarding norms of interaction with the people they perceive as most influential regarding their teaching. Participants also tend to admire the teaching of the people they perceive as most influential. About 95% agreed that they admire their most influential person as a teacher. For the second most influential people the percentage of participants who offered positive responses tended to drop by about 5%. Although still quite positive, there is a difference in the percentage of positive norms indicated for the second-most influential people.

The results of this study give some clues as to how participants perceive faculty to influence their beliefs. It does not appear that faculty influence via teaching observations; less than half of participants indicated that faculty have observed them performing their teaching and provided feedback designed to help them improve. This low percentage is consistent with other studies that have investigated graduate teaching assistant support for teaching (e.g. Justice et al., in press), which have also found low percentages of GTAs who have experienced teaching observations. This result is unfortunate, considering that observations with feedback have been found to be one of the most empirically supported professional development experience (e.g., Williams, 1991).

As suggested by Rumsey (1998), it is possible that faculty influence graduate students' beliefs by participating in required meetings. Participants in this study estimate that faculty attend about 80% of required meetings, on average. On the other hand, it does not appear that faculty spend much time participating in voluntary meetings. Only about 15% of participants indicated that faculty attended voluntary meetings at least once a month, and very few participants (less than 1%) indicate that faculty attended voluntary meetings weekly or more.

5.3 The Relationship Between Statistics Graduate Students' Experiences in CoPs and Beliefs about Teaching Statistics

When using Prior Beliefs as a covariate, results did not indicate much relationship between Current Beliefs and the four constructs designed to measure graduate students' experiences of CoPs. In the theoretical model given in Figure 1 of Chapter 2, most of the paths were eliminated; only the path from prior beliefs to current beliefs was retained. When an ordinary least-squares regression approach was used to analyze four candidate models using two-way interactions of core constructs, the only construct included in the best model was prior beliefs. The best linear model used none of the candidate interactions relating the other core constructs to current beliefs.

The results of this study do not imply that there is no relationship between participation in CoPs and current beliefs about teaching statistics. Rather, this study was not able to detect any relationships using the measures of the six core constructs that were chosen. It is possible that relationships would be detected had the measures of the core constructs been defined and measured differently. Or, it is possible relationships were not detected because two-way interactions using Faculty Presence, Engagement, and Orientation were included in candidate models. As noted by Keith (2006), in social science research interactions can be difficult to detect, often because of measurement error, small sample sizes, or because they simply do not exist.

It is also possible that other models that were not included in this study may capture relationships between participation in CoPs and current beliefs. In the spirit of Keith (2006), this study avoided conducting a "fishing expedition." Interactions of the six core constructs with each of the nine characteristics variables were not included as candidates. Higher-order interactions among the core constructs were not included either. Although some higher-order or characteristic variable interactions are theoretically plausible, there was not enough theory to distinguish a reasonably small set as candidates. It is also possible that the core constructs may have been useful as main effects, however, theory did not point to such model candidates.

Although prior beliefs was the only ***core*** construct in a plausible model, many characteristics of graduate students appear to have potential for explaining variation in statistics graduate students' beliefs about teaching statistics. The model selection methods did not distinguish between the base model (which uses only prior beliefs as a predictor) and models that added main effects for variables such as type of department (biostatistics or statistics), level of degree (master's versus Ph.D.), interest in teaching, international student status, year in program, prior K–12 and college-level teaching experience, and various levels of experience with teaching responsibilities in current degree programs.

These results are consistent with prior research that suggests graduate students with these characteristics have different teaching beliefs, teaching practices, and student experiences (e.g., Boman, 2013; Shannon et al., 1998).

Finally, it is important to note that the chosen models were different when the few (n=8) mathematics graduate students were included in the sample. Although the number of mathematics students was small, their influence on the model was large. This influence could be due to sampling error, or it could also reflect fundamental differences between mathematics and statistics graduate students' experiences. It is possible that graduate student professional development opportunities designed for mathematics or statistics graduate students may not appropriately serve the other group.

5.4 The *Graduate Students' Experiences Teaching Statistics* (GETS) Inventory

The instrument that was developed and used for this study, the *Graduate Students' Experiences Teaching Statistics* (GETS) *Inventory*, was designed to gather statistics graduate students' beliefs about teaching statistics and experiences in their statistics-teaching CoPs. As with other scales that have been developed to measure statistics teaching practices (e.g., Hassad, 2011) and beliefs (e.g., Zieffler et al., 2012), the GETS inventory faced many challenges. Some of the items that were meant to measure the same construct did not have strong or positive correlation coefficients (e.g., Items 27 and 37). The negative and weak correlations indicate that the items do not measure the same construct. Other items may be improved by using different terminology. For example, for Items 37 and 38, the term *focused* may be too vague. Also, participants may have interpreted the term *frequently* (Item 58) in different ways.

Items designed to measure prior beliefs gave some participants particular difficulty. Several participants (n=27) could not provide enough information to be assigned a score for prior beliefs. These participants left blank at least three out of the five items used for the measure. A possible solution might be to offer a limited number of discrete percentage response options (e.g., 0, 20, 40, 60, 80, 100) instead of leaving the item open for participants to enter any quantity between 0 and 100 (inclusive). The discrete response options may reduce cognitive load and give participants some structure to roughly estimate their answers instead of needing to make decisions about minor values.

Items related to topics that have been used in previous instruments (e.g., use of lectures, use of activities) were refined based on suggestions offered by previous researchers. For example, as suggested by Justice et al. (in press), the term *activities* was not used for items about the use of small-group activities. This term was avoided so as to prevent potential confusion with demonstrations conducted in front of the classroom (which are not the same as student-centered activities). Instead, the item used the phrase, "time for students to communicate their ideas together in small groups." Also based on the suggestions of Justice et al. (in press), the type of statistics course for which participants were asked to indicate their beliefs about teaching statistics was more clearly defined. The items specified that the hypothetical course was not online, had about 35 students enrolled, was not calculus-based, and held no separate discussion or lab sessions.

Other items designed to measure statistics teaching beliefs and practices were developed based on the researcher's beliefs of what it means to be student-centered. These (novel) items were not based on prior instruments used to measure teaching beliefs and teaching practices. Some of the novel items appeared useful. For example, Items 33, 34, and 62 had high correlations with other items that have been used to measure student-centered teaching beliefs and teaching practices. These items are based on the use of assessments completed in small groups (e.g., group quizzes, assignments, or projects). Another novel topic that produced high correlations was Item 60. This item asks whether students have the opportunity to communicate their ideas using a variety of means and media (e.g., clickers, chalk boards, oral presentations). These topics have potential to be useful in measuring teaching beliefs and teaching practices in the future. They may be even more useful if revised to use more specific terminology (e.g., replace *variety* with a specific quantity).

There were many items in the GETS Inventory that were designed to measure statistics graduate students' experiences of CoPs. Items designed to measures Engagement in the CoP focused on the frequency, duration, and accessibility of the community. There is room for improvement in measuring this construct. The items in the GETS Inventory did not attempt to measure the *quality* or the *nature* of the engagement in the community. These aspects of engagement may be more important than the time, duration, and accessibility that were included in the inventory.

Items regarding Norms of Interaction were not able to detect variation in norms among participants' CoPs. This issue made it difficult to establish a measure of norms of interaction for the study. It is possible that if the response options were not dichotomous (e.g., Likert-type items using four response options) the items may be able to detect more variation in CoP norms. Also, the lack of variation could have been because influential people tend to have positive norms. However, that does not necessarily mean that there are positive norms in the CoP, as a whole. Perhaps there are other items that would better reflect how participants interact with their CoP, overall, instead of just the influential people.

Two sections of the GETS Inventory were designed to gather information about characteristics of statistics graduate students. Some of the characteristics were not as straightforward as one may expect. One example is international student status. If a participant’s family moved to the United States just before the student began to attend their university, questions arise whether they should be considered an international student or a native student. Or, if the participant holds dual citizenship, the participant may have trouble deciding whether they are considered an international student. For the purposes of this survey, it was decided that the item would focus on whether students obtained a VISA to study in their current degree programs (see Item 65). This approach appeared to work well; the international students in think-aloud and pilot studies made no mention of difficulties answering the question. Also, no comments in the final open-ended portion of the survey indicated difficulty interpreting the item. However it is possible that this approach miscategorized some students.

Another characteristic that required careful thought was regarding degree status (e.g., master’s or doctoral student status). While the target population of the study was doctoral students who would go on to become the future statistics professorate, there is some difficulty distinguishing doctoral students from master’s students. Some degree programs consider all first- or second-year students to be master’s students until they pass their qualifying exams. Other programs may consider all their students doctoral students regardless of whether they have passed their qualifying exams. To account for this problem, Item 67 asks, “Do you *intend* to earn a Ph.D. at your current institution?” Using this strategy, it was hoped that the item could distinguish between (terminal) masters students and masters students who have not advanced to the doctoral stage of their degree programs.

A proxy was used for participants’ interest in teaching. Item 68 asks, “Do you expect to teach statistics courses as part of your primary career?” This approach assumes that if someone expects to pursue a career that involves teaching (namely, academia), they will be more interested in learning about teaching. This item was perhaps the least convincing in its ability to serve its purpose. It is reasonable to believe, simply based on face value, that there may be other (better) ways of obtaining information about participants’ interest in teaching statistics.

5.5 Limitations

Many of the limitations of this study are related to difficulties measuring the constructs of interest. For example, the CoP Orientation construct was central to the theoretical models proposed in this study. All but one of the theoretical regression models used an interaction with the Orientation variable. However a simpler construct was used in place of the CoP Orientation. As a proxy for the CoP Orientation, items were included regarding the teaching practices of the department member participants perceived to be most influential. The practices of one person may not reflect the orientation of the community, however this proxy was used to keep the survey a reasonable length. In addition, fit measures for the CoP Orientation construct were not adequate. The study was not able to adequately measure this very central construct.

Another construct that was difficult to measure was the Norms of Interaction. For this sample the norms were predominantly positive, and there was not enough variation in the items to establish a measure for Norms. As a result this study was not able to discern how Norms of Interaction play a part in the effects of participation in a CoP upon a graduate student's teaching beliefs.

There are also limitations due to unmet assumptions for the methods of analyses. For example, the ceiling effect for the outcome measure of Current Beliefs infringes upon the assumption that residuals are normally distributed (see Figure 10, Chapter 4). Also, it may have been more appropriate to use Item Factor Analysis instead of Confirmatory Factor Analysis for determining the Norms and Orientation core construct measures, because the items for these constructs were not based on continuous response options. There were also issues with assumptions for the cross-validation procedures used. According to Arlot and Celisse (2010), cross-validation techniques protect against overfitting when the training sample is independent from the validation sample and the data are independently distributed. However, the clustered nature of the data (students are nested within schools) calls this assumption into question.

Some other methodological limitations arise from the relatively small sample size for the study. According to Burnham and Anderson (2004), AIC is an approximation for the K-L information when working with large samples (and good models). In the case of this study, the small sample sizes may require even more theoretical justification for the models in order to use the AIC. Also, the small sample size may have led to insignificant results due to lack of power. One reason for the small sample size may be that that another graduate student survey was sent to the statistics education community at about the same time that the GETs Inventory invitation was sent. The other survey, which also offered a gift card in exchange for participation, may have competed for graduate students' time and energy. Or, students may have confused the surveys, thinking that they were the same.

As mentioned previously, the sample was not randomly selected and should not be generalized to the population of all statistics graduate students in the United States. While measures were taken to try to recruit participants from a variety of institutions (e.g., offering participation in random drawings for five $25 Amazon.com gift cards), graduate students who are more interested in teaching may be more inclined to take time to complete the survey. As with all voluntary response sampling methods, it is also plausible that the views are largely polarized. Graduate students who have had particularly positive or, perhaps more likely, particularly negative experiences may be more willing to participate as an outlet for sharing their experiences.

5.6 Implications for Future Research.

This study was the first to examine relationships between statistics graduate students' beliefs about teaching and their participation in CoPs. Some of the challenges faced in this study illuminate future studies that may contribute to a better understanding of graduate students' experiences with professional development related to teaching.

5.6.1 The role of faculty. While this study suggests that faculty are often perceived to be the most influential department member, there are still open questions about how they may influence graduate students' teaching beliefs, and to what extent. As potential examples, faculty may influence graduate students through interactions in required weekly meetings, in casual conversations, or by demonstrating particular teaching practices in the courses they teach. Future research could also be conducted to explore the roles of the people who are perceived to have the most influence. They may be lead instructors for courses that graduate students are hired to teach or assist, teaching mentors, or they may be graduate students' research advisers.

5.6.2 The role of positive norms of interaction. This study was not able to use the Norms measure in candidate models because there was not enough variation in norms scores. Most of the norms were positive. This result raises questions about whether there is a relationship between positive interactions and potential to influence beliefs. Future studies could be designed to measure norms of interaction, and to explore whether those who are able to influence teaching beliefs tend to have positive norms of interaction. Research in this area may point to aspects of relationships that can help and hinder changes in beliefs about teaching.

5.6.3 Measuring interest in teaching. Interest in teaching is a theoretically compelling characteristic of graduate students that may affect teaching beliefs and their participation in CoPs. There is a need for more research investigating ways to measure graduate students' interest in teaching. For this study, a proxy was used: (Item 68) regarding participants' expectation to teach as a part of their primary career. There are other aspects of interest in teaching that could be useful in measuring interest in teaching statistics. However, the difficulty in creating items and measuring interest in teaching should not be underestimated, as inventories have been written to try to measure related topics (e.g., attitudes toward statistics, SATS, Schau et al., 1995).

5.6.4 Measuring experiences of CoPs. One of the greatest challenges of this study was the attempt to measure participants' experiences of CoPs. These challenges are not surprising because previous literature suggests CoPs are often difficult to identify and may occur only informally. Still, there is a great need for research that can identify methods of measuring and detecting how communities form, what defines them, and characteristics that affect their influence upon participants. It may be appropriate to start with careful observations of CoPs to learn of more ways that CoPs are embodied. Studies could also look for other constructs that can be used to identify important characteristics of participants' experiences of them. Qualitative methods could be used to establish theoretical models relating interactions of CoPs characteristics and participant characteristics.

For the constructs that were identified to characterize CoPs in this study, future research could be conducted to learn how to measure them. For example, to gather data regarding participants' Engagement, researchers could expand beyond frequency and time to also include aspects such as intellectual engagement (e.g., interest in the topics discussed). Also, in this study the CoP orientation toward student-centered teaching construct was substituted by a proxy (the most influential department member's orientation toward student-centered teaching). Further research could be conducted to more appropriately measure the beliefs of the entire community rather than just one person.

5.6.5 Measuring teaching beliefs and teaching practices. While the GETS Inventory was able to build upon previous instruments that are designed to measure beliefs about teaching statistics and teaching practices (e.g., GSSTI, Justice et al., in press), there is still much need for further research that can help to measure teaching beliefs and practices. Items in the GETS Inventory that were not used in the calculation of the measures of constructs should be examined, possibly removed or revised, piloted, and evaluated again. Also, more validity evidence should be collected (e.g., data from interviews, relationships with other instruments designed to measure teaching beliefs) to support the intended uses of the GETS Inventory in assessing graduate students' teaching beliefs. Instruments that have been used for this purpose in the past (e.g., the Reformed Teaching Observation Protocol, Sawada et al., 2002) may need to be revised to account for the limited autonomy that most graduate students have in the classroom.

Some of the novel items regarding teaching beliefs and teaching practices on the GETS Inventory that showed promise might be examined in future studies. For example, the use of group assessments, a topic that has not been included in previous instruments designed to measure statistics teaching beliefs and teaching practices, had mild theoretical and empirical basis for being used to measure teaching beliefs. This item could be included in future instruments and CFA analyses could be conducted to gather more empirical basis for its use measuring student-centered teaching beliefs. Also, based on CFA results and theory, the item regarding students communicating their ideas using a variety of means and media (e.g., clickers, chalk boards, oral presentations) was used measure orientation toward student-centered teaching. Further research could include think-aloud and pilot studies that further explore whether this item can help measure teaching practices.

5.7 Conclusion

This study explored graduate students' experiences in CoPs and their teaching beliefs. The survey developed for the study offers items that were useful in measuring graduate student characteristics, their teaching beliefs, and their experiences in CoPs. Results indicated that most participants have experienced some forms of CoPs related to teaching statistics. Participants' teaching beliefs have typically become more student-centered since entering their degree programs, although often not by much. Results also indicated that faculty tend to be selected as the department member with the largest influence upon participants' teaching beliefs. The department members who are perceived to influence graduate students' teaching beliefs typically are viewed as people who acknowledge graduate students' ideas and who maintain respect even when they disagree.

For some of the constructs used to describe graduate students' perceived experiences of CoPs the data did not result in adequate measures (e.g., Norms of Interaction, Orientation). This issue made it difficult to examine relationships between these aspects of CoPs or interactions that may involve them. However, this study laid a foundation for future research that aims to investigate questions related to graduate students' beliefs about teaching statistics, their experiences of CoPs, and their professional development related to teaching.

References

Abbott, R. D., Wulff, D. H., & Szego, C. K. (1989). Review of research on TA training. *New Directions for Teaching and Learning*, (39), 111–124. doi:10.1002/tl.37219893912

American Association for the Advancement of Science (AAAS). (1989). *Science for all Americans: A project 2061 report on literacy goals in science, mathematics, and technology*. Washington, D.C.: AAAS.

American Statistical Association. (2005). *GAISE college report*. http://www.amstat.org/education/gaise/GAISECollege.htm

American Statistical Association Revision Committee. (2016). *GAISE College Report 2016*. http://www.amstat.org/education/gaise.

Anderson, E. M., & Shannon, A. L. (1988). Toward a conceptualization of mentoring. *Journal of Teacher Education, 39*(1), 38–42.

Arbaugh, J. B., Cleveland-Innes, M., Diaz, S. R., Garrison, D. R., Ice, P., Richardson, J. C., & Swan, K. P. (2008). Developing a community of inquiry instrument: Testing a measure of the Community of Inquiry framework using a multi-institutional sample. *Internet and Higher Education*, *11*(3–4), 133–136. doi:10.1016/j.iheduc.2008.06.003

Arlot, S. & Celisse, A. (2010) A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4, 40–79.

Austin, A. E. (2002). Preparing the next generation of faculty: Graduate school as socialization to the academic career. *The Journal of Higher Education*, 73(1), 94–122.

Bailey, K. (1983). Foreign Teaching Assistants at U.S. Universities: Problems in Interaction and Communication. *TESOL Quarterly*, *17*(2), 308–310.

Bassili, J. N., & Scott, B. S. (1996). Response latency as a signal to question problems in survey research. *Public opinion quarterly*, *60*(3), 390–399.

Beaujean, A. A. (2014). *Latent variable modeling using R: A step-by-step guide*. New York: Routledge.

Belnap, J. K. (2005). *Putting TAs into Context : Understanding the Graduate Mathematics Teaching Assistant*. The University of Arizona.

Birch, J. B., & Morgan, J. (2005). TA training at Virginia Tech: A stepwise progression. *American Statistician*, *59*(1), 14–18.

Blair, R., Kirkman, E. & Maxwell, J. (2013). *Statistical abstract of undergraduate programs in the mathematical sciences in the United States: Fall 2010 CBMS Survey*. Washington, DC: American Mathematical Society.

Blumenfeld, P. C., Fishman, B. J., Krajcik, J. S., and Marx, R. W. 2000. Creating usable innovations in systemic reform: Scaling up technology-embedded project-based science in urban schools. *Educational Psychologist 35*(3):149–164.

Boehrer, J., & Sarkisian, E. (1985). The Teaching Assistant's Point of View. New *Directions for Teaching and Learning*, (22), 7–20.

Boice, R. (1992). Lessons learned about mentoring. *New Directions for Teaching and Learning, 50*, 51–61.

Boman, J. S. (2013). Graduate student teaching development : Evaluating the effectiveness of training in relation to graduate student characteristics. *Canadian Journal of Higher Education*, *43*(1), 100–114.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3–15.

Bray, J. H., & Howard, G. S. (1980). Methodological considerations in the evaluation of a teacher-training program. *Journal of Educational Psychology*, *72*(1), 62–70. doi:10.1037//0022-0663.72.1.62

Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science, 16*(3), 199–231.

Buerkel-Rothfuss, N., & Gray, P. (1990). Graduate teaching assistant training in speech communication and non-communication departments: A national survey. *Communication Education*, *39*(4), 292–307. Retrieved from http://www.tandfonline.com/doi/full/10.1080/03634529009378811

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261–304.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology, 65*(1), 23–35.

Byrd, P., Constantinides, J. C., & Pennington, M. (1989). *The Foreign Teaching Assistant's Manual*. New York: Collier MacMillan.

Carroll, J. G. (1980). Effects of training programs for university teaching assistants: A review of empirical research. *The Journal of Higher Education*, *51*(2), 167. doi:10.2307/1981372

Cahyadi, V., & Butler, P. H. (2005). A professional development course for teaching assistants. In *HERDSA Annual International Conference* (pp. 79–85).

Chism, N. V. N. (Ed.). (1987). *Institutional responsibilities and responses in the employment and education of teaching assistants: Readings from a national conference*. Columbus, Ohio: The Ohio State University Center for Teaching Excellence.

Cobb, G. (1992). Teaching Statistics. Steen, L.A. (Ed.) *Heeding the call for change: Suggestions for curricular action*. MAA Notes (22). Washington, DC: Mathematical Association of America, 3–34.

Constantinides, J. C. (1987). Designing a training program for international teaching assistants. In N. V. N. Chism (Ed.), *Institutional responsibilities and responses in the employment and education of teaching assistants: Readings from a national conference.* (pp. 275–283). Columbus, Ohio: The Ohio State University Center for Teaching Excellence.

Constantinides, J. C. (1989). ITA training programs. *New Directions for Teaching and Learning*, (39), 71–77. doi:10.1002/tl.37219893908

Cooney, T. J. (1994). On the application of science to teaching and teacher education. In R. Biehler, R. W. Scholz, R. Sträer, & B. Winkelmann (Eds.), Didactics of mathematics as a scientific discipline (pp. 103–116). Dordrecht, The Netherlands: Kluwer.

Cox, M. D. (2004). Introduction to faculty learning communities. *New directions for teaching and learning*, *2004*(97), 5–23.

Crede, E., BorrEgo, M., & McNair, L. D. (2010). Application of community of practice theory to the preparation of engineering graduate students for faculty careers. *Advances in Engineering Education*, *2*(2), 1–22.

Dalgaard, K. A. (1982). Some effects of training on teaching effectiveness of untrained university teaching assistants. *Research in Higher Education*, *17*(1), 39–50.

Darling, A. L. (1987). TA socialization: A communication perspective. In N. V. N. Chism (Ed.), *Preparing Graduate Students to Teach: Past, Present, and Future. The professional development of graduate teaching assistants, 1–18.* (pp. 91–94). Columbus, Ohio: The Ohio State University.

Darling, A., & Dewey, M. (1990). Teaching assistant socialization: Communication with peer leaders about teaching and learning. *Teaching and Teacher Education*, *6*(4), 315 – 326. doi:10.1016/0742-051X(90)90024-Y

Darling, A., & Staton, A. (1989). Socialization of graduate teaching assistants: A case study in an American university. *International Journal of Qualitative Studies in Qualitative Education*, *2*(3), 221 – 235.

Davis, T. M. (2000). *Open doors: Report on international education exchange.* (pp. 1–101). New York, NY.

Davis, W. E. (1991). International teaching assistants and cultural differences: Student evaluations of rapport, approachability, enthusiasm and fairness. In *Preparing the Professoriate of Tomorrow to Teach. Selected Readings in TA Training.* (pp. 446–451). Dubuque, IA: Kendall/Hunt Publishing Company.

Davis, W. E., & Minnis, D. L. (1993). Designing a program to prepare graduate students for careers as college teachers. *Innovative Higher Education*, *17*(3), 211–224. doi:10.1007/BF00915602

DeNeef, A. L. (2002). *The Preparing Future Faculty Program: What difference does it make?* (pp. 1–31). Washington, DC.

Diamond, R. M., & Gray, P. J. (1987). A national study of teaching assistants. In *Association for the Study of Higher Education* (pp. 1–9).

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method*. *Internet Mail and Mixed Mode Surveys The tailored design method* (3rd ed.).

Dodge, L., & Kendall, M. E. (2004). Learning communities. *College Teaching*, *52*(4), 150–155.

Dotger, S. (2011). Exploring and developing graduate teaching assistants' pedagogies via lesson study. *Teaching in Higher Education*, *16*(2), 157–169. doi:10.1080/13562517.2010.507304

Ebel, R. L. (1954). Procedures for the analysis of classroom tests. *Educational and Psychological Measurement*, *14*(2), 352–364.

Ebert-May, D., Derting, T. L., Hodder, J., Momsen, J. L., Long, T. M., & Jardeleza, S. E. (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

Ethington, C. A., & Pisani, A. (1993). The RA and TA experience: Impediments and benefits to graduate study. *Research in Higher Education*, *34*(3), 343–354.

Fagen, A. P., & Suedkamp Wells, K. M. (2004). The 2000 national doctoral program survey: An on-line study of students' voices. In D. H. Wulff & A. E. Austin (Eds.), *Paths to the professoriate: Strategies for enriching the preparation of future faculty* (pp. 74–91).

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research, 38*(1), 47–65.

Farbrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. doi:10.1037/1082-989X.4.3.272

Fernald, P. S. (1995). Preparing psychology graduate students for the professoriate. *American Psychologist, 50*(6), 421.

Fisher, M. (1985). Rethinking the "foreign TA problem." *New Directions for Teaching and Learning*, (22), 63–73.

Ford, J. E., Gappa, L., Wendorff, J., & Wright, D. L. (1991). Model of an ITA Institute. In *Preparing the Professoriate of Tomorrow to Teach. Selected Readings in TA Training.* (pp. 341–355). Dubuque, IA: Kendall/Hunt Publishing Company.

Fox, J. (2016). Polycor: Polychoric and Polyserial Correlations. R package version 0.7-9.

Frank, B. W., & Speer, N. M. (2013). Building knowledge for teaching : Three cases of physics graduate students. In *AIP Conference Proceedings, 1513* (Vol. 1513, p. 126).

Froelich, A. G., Duckworth, W. M., & Stephenson, W. R. (2005). Training Statistics teachers at Iowa State University. *The American Statistician*, *59*(1), 8–10.

Garfield, J., & Everson, M. (2009). Preparing teachers of statistics : A graduate course for future teachers. *Journal of Statistics Education*, *17*(2), 223 – 237.

Garfield, J., Hogg, R., Schau, C., & Wittinghill, D. (2002). First courses in statistical science: The status of educational reform efforts. *Journal of Statistics Education, 10* (2).

Gelman, A. (2005). A course on teaching statistics at the university level. *The American Statistician*, *59*(1), 4–7.

Garfield, J., & Everson, M. (2009). Preparing teachers of statistics : A graduate course for future teachers. *Journal of Statistics Education*, *17*(2), 223 – 237.

Garrison, D., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, *2*(2), 87–105.

Garrison, D. R., Anderson, T., & Archer, W. (2010). The first decade of the community of inquiry framework: A retrospective. *Internet and Higher Education*, *13*(1–2), 5–9. doi:10.1016/j.iheduc.2009.10.003

Gilmore, J. A., Maher, M. A., Feldon, D. F., & Timmerman, B. E. (2013). Exploration of factors related to the development of science, technology, engineering and mathematics gradate teaching assistants' teaching orientations. *Studies in Higher Education*, 1–19.

Gilreath, J. A., & Slater, T. F. (1994). Training graduate teaching assistants to be better undergraduate physics educators. *Physics Education*, *29*(4), 200–203. doi:10.1088/0031-9120/29/4/003

Golde, C. M., & Dore, T. M. (2001). *At cross purposes: What the experiences of today's doctoral students reveal about doctoral education* (pp. 1–65).

Green, J. L. (2010). Teaching highs and lows: Exploring university teaching assistants' experiences. *Statistics Education Research Journal*, *9*(2), 108–122.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology Journal, 29*(2), 75.

Hammrich, P. L. (1996). The Impact of teaching assistants' conceptions on college science teaching. *The journal of Graduate Teaching Assistant Development*, *3*, 109–117.

Hannigan, A., Gill, O., & Leavy, A. M. (2013). An investigation of prospective secondary mathematics teachers' conceptual knowledge of and attitudes towards statistics. *Journal of Mathematics Teacher Education, 16*(6), 427–449.

Harkness, W. L., & Rosenberger, J. L. (2005). Training graduate students at Penn State University in teaching Statistics. *The American Statistician*, *51*(1), 11–13.

Hartnett, R. T., & Katz, J. (1977). The education of graduate students. *Journal of Higher Education*, *48*(6), 646–664. doi:10.2307/1979010

Hassad, R. A. (2011). Constructivist and behaviorist approaches: Development and initial evaluation of a teaching practice scale for introductory statistics at the college level. *Numeracy*, *4*(2). doi:10.5038/1936-4660.4.2.7

Healy, C. C., & Welchert, A. J. (1990). Mentoring relations: A definition to advance research and practice. *Educational Researcher*, *19*(9), 17–21.

Hoekje, B., & Williams, J. (1992). Communicative competence and the dilemma of international teaching assistant education. *TESOL Quarterly*, *26*(2), 243–269.

Hoessler, C., & Godden, L. (2015). The visioning of policy and the hope of implementation: Support for graduate students' teaching at a Canadian institution. *The Canadian Journal of Higher Education, 45*(1), 83.

Hogg, R. V. (1991). Statistical education : Improvements are badly needed. *The American Statistician*, *45*(4), 342–343.

Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity, 44*(1), 153-166.

Holmes, N. G., Ives, J., & Warren, M. (2013). Teaching assistant professional development by and for TAs. *The Physics Teacher*, *51*(4), 218–219.

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.

Jacobs, L. C., & Friedman, C. B. (1988). Student achievement under foreign teaching associates compared with native teaching associates. *The Journal of Higher Education*, 551–563.

Jones, J. L. (1993). TA training: From the TA's point of view. *Innovative Higher Education*, *18*(2), 147–161.

Justice, N., Zieffler, A., & Garfield, J. (in press) Statistics Graduate Teaching Assistants' Beliefs, Practices, and Preparation for Teaching Introductory Statistics. *Statistics Education Research Journal.*

Kane, R., Sandretto, S., & Heath, C. (2002). Telling half the story: A critical review of research on the teaching beliefs and practices of university academics. *Review of Educational Research*, *72*(2), 177–228. doi:10.3102/00346543072002177

Keith, T. Z. (2006). *Multiple regression and beyond*. Boston: Pearson Education, Inc.

Kember, D. (1997). A reconceptualisation of the research into university academics' conceptions of teaching. *Learning and Instruction, 7*(3), 255–275.

Korpan, C. (2014). From teaching assistant training (TA) to workplace learning. *Collected Essays on Learning and Teaching, 7*(2). Retrieved from http://celt.uwindsor.ca/ojs/leddy/index.php/CELT/article/viewFile/4002/3294

Lave, J. (1996). Teaching, as learning, in practice. *Mind, Culture, and Activity, 3*(3), 149–164.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, England: Cambridge University Press.

Little, J. W. (1990). The mentor phenomenon and the social organization of teaching. *Review of Research in Education*, *16*, 297–351.

Little, J.W. (2002). Locating learning in teachers' communities of practice: Opening up problems of analysis in records of everyday work. *Teaching and Teacher Education, 18*(8), 917–946.

Luft, J. A., Kurdziel, J. P., Roehrig, G. H., & Turner, J. (2004). Growing a garden without water : Graduate teaching assistants in introductory TA training at a doctoral research university. *Journal of Research in Science Teaching*, *41*(3), 211–233. doi:10.1002/tea.20004

Luo, J., Grady, M. L., & Bellows, L. H. (2001). Instructional issues for teaching assistants. *Innovative Higher Education*, *25*(3), 209–230.

Marincovich, M., Prostko, J., & Stout, F. (Eds.). (1998). *The professional development of graduate teaching assistants*. Bolton, MA: Anker Publishing Company.

Martins, J. A., Nascimento, M. M., & Estrada, A. (2012). Looking back over their shoulders: A qualitative analysis of Portuguese teachers' attitudes towards statistics. *Statistics Education Research Journal, 11*(2), 26–44.

McKimm, J., Jollie, C., & Hatter, M. (2007). Mentoring: Theory and practice. London Deanery Faculty Development: London.

McLaughlin, M. W., & Mitra, D. 2001. Theory-based change and change-based theory: Going deeper, going broader. *Journal of Educational Change 1*(2):1–24.

Moore, D.S. (2005). Preparing graduate students to teach statistics: Introduction. *The American Statistician*, *59*(1), 1–3.

Muzaka, V. (2009). The niche of graduate teaching assistants (GTAs): perceptions and reflections. *Teaching in Higher Education*, *14*(1), 1–12. doi:10.1080/13562510802602400

Myers, S. A. (1994). The availability and helpfulness of graduate teaching assistant socialization activities. *Communication Research Reports*, *11*(2), 221–228.

Myers, S. A. (1998). GTAs as organizational newcomers: The association between supportive communication relationships and information seeking. *Western Journal* *of Communication*, *62*(1), 54–73. Retrieved from http://www.tandfonline.com/doi/full/10.1080/10570319809374597

Noll, J. A. (2011). Graduate teaching assistants' statistical content knowledge of sampling. *Statistics Education Research Journal*, *10*(2), 48–74.

Nyquist, J. D., Abbott, R. D., Wulff, D. H., & Sprague, J. (Eds.). (1991). *Preparing the professoriate of tomorrow to teach: selected readings in TA training*. Dubuque, Iowa: Kendall/Hunt Publishing Company.

Nyquist, J. D., & Sprague, J. (1998). Thinking developmentally about TAs. In *The professional development of graduate teaching assistants* (pp. 61 – 68).

Nyquist, J. D., & Wulff, D. H. (1987). The training of graduate teaching assistants at the University of Washington. In N. V. N. Chism (Ed.), *Institutional responsibilities and responses in the employment and education of teaching assistants: Readings from a national conference.* (pp. 144–154). Columbus, Ohio: The Ohio State University Center for Teaching Excellence.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, *62*, 307–332. doi:10.3102/00346543062003307

Park, C., & Ramos, M. (2002). The donkey in the department? Insights into the graduate teaching assistant (GTA) experience in the UK. Journal of Graduate Education, 3, 47–53.

Parrett, J. L. (1987). A ten-year review of TA training programs: Trends, patterns, and common practices. In N. V. N. Chism (Ed.), *Institutional responsibilities and responses in the employment and education of teaching assistants: Readings from a national conference.* (pp. 67–79). Columbus, Ohio: The Ohio State University Center for Teaching Excellence.

Prieto, L. R., & Altmaier, E. M. (1994). The relationship of prior training and previous teaching experience to self-efficacy among graduate teaching assistants. *Research in Higher Education*, *35*(4), 481 – 497.

Prieto, L. R., & Meyers, S. A. (1999). Effects of training and supervision on the self-efficacy of Psychology graduate teaching assistants. *Teaching of Psychology*, *26*(4), 264–266.

Pruitt-Logan, A. S., & Gaff, J. G. (2004). Preparing Future Faculty: Changing the culture of doctoral education. In D. H. Wulff & A. E. Austin (Eds.), *Paths to the professoriate: Strategies for enriching the preparation of future faculty* (pp. 177–193). San Francisco, CA: Jossey-Bass.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher, 29*(1), 4–15.

Qualtrics. (2015). Qualtrics. Provo, Utah.

R Development Core Team. (2014). R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria.

Richlin, L., & Cox, M. D. (2004). Developing scholarly teaching and the scholarship of teaching and learning through faculty learning communities. *New directions for teaching and learning*, *2004*(97), 127–135.

Rodriques, R. A. B., & Bond-Robinson, J. (2006). Comparing faculty and student perspectives of graduate teaching assistants' teaching. *Journal of Chemical Education*, 83(2), 305–312.

Roehrig, G. H., Luft, J. A., Kurdziel, J. P., & Turner, J. A. (2003). Graduate teaching assistants and inquiry-based instruction : Implications for graduate teaching assistant training. *Journal of Chemical Education*, *80*(10), 1206 – 1210.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. Retrieved from http://www.doaj.org/doaj?func=fulltext&aId=1325391\nhttp://www.jstatsoft.org/v48/i02

Rovai, A. P. (2002). Development of an instrument to measure classroom community. *Internet and Higher Education*, *5*(3), 197–211. doi:10.1016/S1096-7516(02)00102-1

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, *33*(4), 511–531. doi:10.1007/BF00973770

Rumsey, D. J. (1998). A cooperative teaching approach to introductory statistics. *Journal of Statistics Education*, *6*(1), 163–171.

Sarkisian, E., & Maurer, V. (1998). International TA training and beyond: Out of the program and into the classroom. In *The Professional Development of Graduate Teaching Assistants* (pp. 163–180). Bolton, MA: Anker Publishing Company.

Saroyan, A., Dagenais, J., & Zhou, Y. (2008). Graduate students' conceptions of university teaching and learning: formation for change. *Instructional Science*, *37*(6), 579–600. doi:10.1007/s11251-008-9071-8

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics*, *102*(6), 245–253.

Schau, C., Stevens, J., Dauphinee, T. L., & Vecchio, A. D. (1995). The Development and Validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement, 55*(5), 868–875. doi:10.1177/0013164495055005022

Schlager, M. S., & Fusco, J. (2003). Teacher professional development, technology, and communities of practice: Are we putting the cart before the horse?. *The Information Society*, *19*(3), 203–220.

Schoem, D., Carlton, S., Gates, B., & Black, B. (1991). Developing and implementing a college-wide TA training program at the University of Michigan. In J. D. Nyquist, R. D. Abbott, D. H. Wulff, & J. Sprague (Eds.), *Preparing the Professoriate of Tomorrow to Teach. Selected Readings in TA Training.* (pp. 150–156).

Shannon, D. M., Twale, D. J., & Moore, M. S. (1998). TA teaching effectiveness: The impact of training and teacher experience. *Journal of Higher Education*, *69*(4), 440–466.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information, 22*(2), 63-75.

Simmons, P. E., Emory, A., Carter, T., Coker, T., Finnegan, B., Crockett, D., Richardson, L., Yager, R., Craven, J., Tillotson, J., Brunkhorst, H., Twiest, M., Hossain, K., Gallagher, J., Duggan-Haas, D., Parker, J., Cajas, F., Alshannag, Q., McGlamery, S., Krockover, J., Adams, P., Spector, B., LaPorta, T., James, B., Rearden, K., & Labuda, K. (1999). Beginning teachers: Beliefs and classroom actions. *Journal of Research in Science Teaching, 36*(8), 930–954.

Smylie, M. A., Allensworth, E., Greenberg, R. C., Harris, R., & Luppescu, S. (2001). *Teacher professional development in Chicago: Supporting effective practice*. Chicago, IL: Consortium on Chicago School Research.

Sowder, J. T. (2007). The mathematical education and development of teachers. In F. K. Lester (Ed.), *Second Handbook of Research on Mathematical Thinking and Learning.* Charlotte, NC: Information Age Publishing.

Speer, N. M. (2004). Meeting the changing needs of mathematics graduate student teaching assistants: The evolution of professional development activities for instruction using collaborative groups. *Journal of Faculty Development*, *20*(1), 5–10.

Staton, A. Q., & Darling, A. L. (1989). Socialization of teaching assistants. *New Directions for Teaching and Learning*, (39), 15–22. doi:10.1002/tl.37219893904

Svinicki, M. D. (1989). The Development of TAs : Preparing for the future while enhancing the present. *New Directions for Teaching and Learning*, *1989*(37), 71–80.

Thomas, C. F., & Monoson, P. K. (1993). Oral English language proficiency of ITAs: Policy, implementation, and contributing factors. *Innovative Higher Education*, *17*(3), 195–209. doi:10.1007/BF00915601

Travers, P. D. (1989). Better teaching training assistants. *College Teaching*, *37*(4), 147–149.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, *24*(1), 80–91.

Volkmann, M. J., & Zgagacz, M. (2004). Learning to teach physics through inquiry: The lived experience of a graduate teaching assistant. *Journal of Research in Science Teaching*, *41*(6), 584–602. doi:10.1002/tea.20017

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods, 17*(2), 228–43.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Weakliem, D. L. (1999). A Critique of the Bayesian Information Criterion for Model Selection. *Sociological Methods & Research, 27*(3), 359–397.

Weimer, M., Svinicki, M. D., & Bauer, G. (1989). Designing programs to prepare TAs to teach. *New Directions for Teaching and Learning*, (39), 57–70. doi:10.1002/tl.37219893907

Wenger, E., McDermott, R. A., & Snyder, W. (2002). *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Harvard, MA: Harvard Business Press.

Wickham, H. (2009). ggplot2: Elegant graphics for data analysis. R package version 2.2.1.

Williams, D. E., & Roach, D. K. (1992). Graduate teaching assistant perceptions of training programs. *Communication Research Reports*, *9*(2), 183–192.

Williams, L. S. (1991). The effects of a comprehensive teaching assistant training program on teaching anxiety and effectiveness. *Research in Higher Education*, *32*(5), 585–598. doi:10.1007/BF00992630

Wise, S. L. (1985). The development and validation of a scale measuring attitudes toward statistics. *Educational and Psychological Measurement, 45*(2), 401-405.

Wulff, D. H., & Austin, A. E. (Eds.). (2004). *Paths to the Professoriate: Strategies for Enriching the Preparation of Future Faculty*. San Francisco, CA: Jossey-Bass.

Wulff, D. H., Austin, A. E., Nyquist, J. D., & Sprague, J. (2004). The development of graduate students as teaching scholars: A four-year longitudinal study. In D. H. Wulff & A. E. Austin (Eds.), *Paths to the professoriate: Strategies for enriching the preparation of future faculty* (pp. 46–73).

Wulff, D. H., Nyquist, J. D., & Abbott, R. D. (1991). Developing a TA program that reflects the culture of the institution: TA training at the University of Washington. In J. D. Nyquist, R. D. Abbott, D. H. Wulff, & J. Sprague (Eds.), *Preparing the Professoriate of Tomorrow to Teach. Selected Readings in TA Training.* (pp. 113–122). Dubuque, Iowa: Kendall/Hunt Publishing Company.

Wyse, S. (2010). *Breaking the mold: Preparing graduate teaching assistants to teach as they are taught to teach*. Michigan State University.

Young, R. (1989). Introduction. *English for Specific Purposes*, *95*(3), 101–107.

Zhang, Y., & Yang, Y. (2015). Cross-validation for selecting a model selection procedure. *Journal of Econometrics, 187*(1), 95–112.

Zieffler, A., Park, J., Garfield, J., DelMas, R., & Bjornsdottir, A. (2012). The Statistics Teaching Inventory: A survey on statistics teachers' classroom practices and beliefs. *Journal of Statistics Education*, *20*(1), 1–29.

APPENDIX A: Theoretical Model Relating Graduate Students' Experiences of CoPs and Their Beliefs about Teaching Statistics

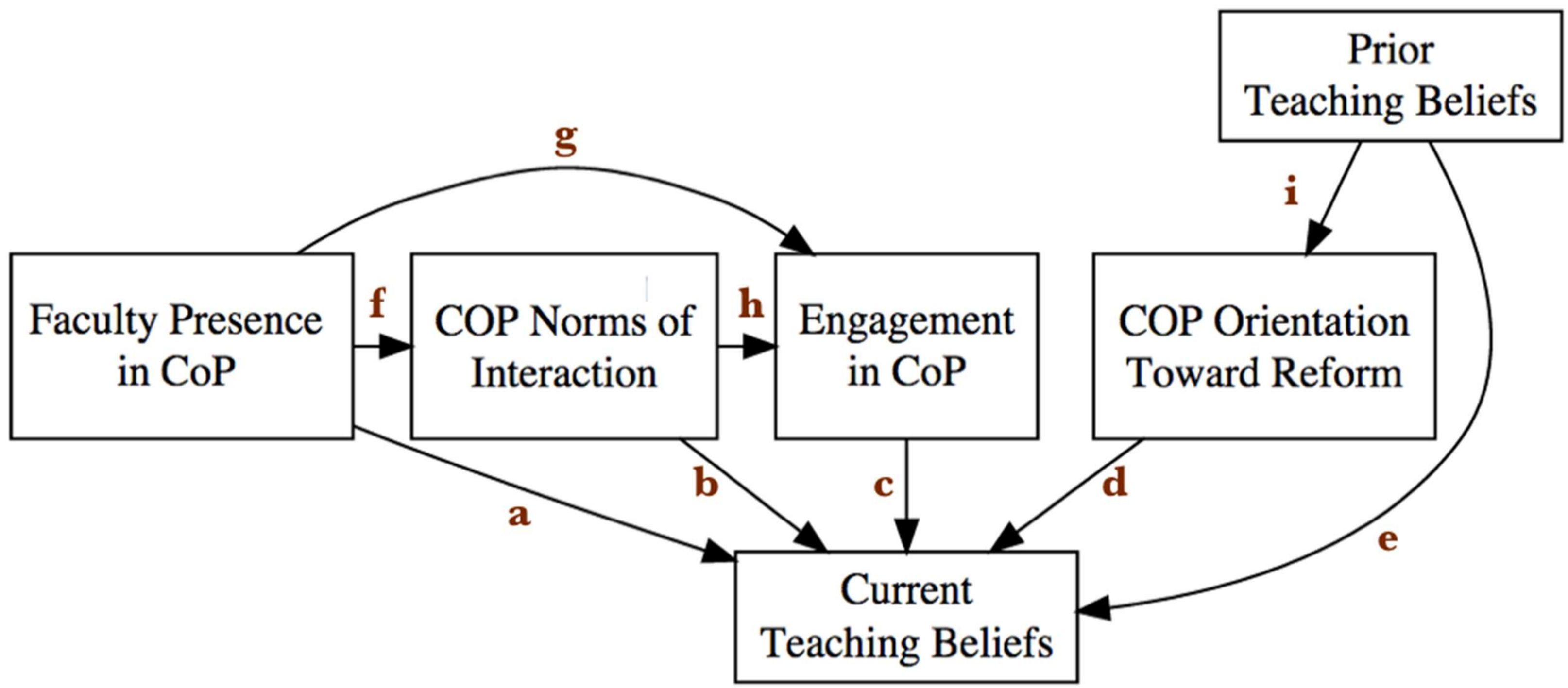


Figure 11. Theoretical model relating four aspects of CoPs to current and prior teaching beliefs with path coefficients. For easier viewing of the relationships in the theoretical model, the disturbances (errors) for endogenous variables are not included in this figure.

APPENDIX B: GETS Inventory Survey Instrument and Response Summaries

(with Response Percentages and Numerical Summaries for Cases Used in Final Model)

Please note that many of these items ought to be adjusted before being used in future instrument. See the section of the Discussion Chapter of this paper that is focused on the items in the GETS Inventory.
Note Also: Item numbers were not included in the original survey; they were later added for ease of reference.

Section 1 (out of 6): You and Your Graduate Program

1. Are you currently enrolled as a graduate student?

Yes	100%
No	0%

If No Is Selected, Then Skip To End of Survey

2. Please enter the number of years you have completed in your current graduate degree program. Please round up (e.g., if you have completed 3.25 years please enter 4).

M	SD	Min	Q1	Med	Q3	Max
2.8	2.5	0	1	2	4	32

3. What best describes the type of department that houses your degree program?

Statistics	70%
Mathematics	0%
Mathematics Education	0%
Biostatistics or Public Health	30%
Business	0%
Computer Sciences	0%
Educational Psychology or Educational Studies	0%
Psychology or Sociology	0%
Other (Please Specify) _______________	0%

At your current institution have you ever been hired in positions that include the following teaching- or research-related activities?

	Yes	No
4. Grading papers	76%	24%
5. Holding office hours or working in a tutorial center	78%	22%
6. Facilitator of a lab or discussion section that meets regularly	59%	41%
7. Primary instructor for a course	32%	68%
8. Assistant to a primary instructor for a course (e.g., attends and helps during class sessions, writes exams).	43%	57%
9. Supervisor of other graduate students for teaching responsibilities	12%	88%
10. Research assistant	57%	43%

Section 2: Interactions With Others in Your Department

11. In your current graduate degree program have you ever had a common or shared space (e.g., office or cubicle) with other graduate students who are involved with the teaching or assisting of a statistics course?

Yes	85 %
No	15%

12. Typically, about how many days per week do you visit your common or shared space (e.g., office or cubicle)?

N	M	SD	Min	Q1	Med	Q3	Max
185	4	1.6	0	3	5	5	7

13. In your current degree program, have you ever been required to meet with other graduate students, faculty, or staff to discuss topics related to teaching or your teaching responsibilities?

Yes	75 %
No	25%

14. On average, about how often have you been required to attend meetings with other graduate students, faculty, or staff to discuss topics related to teaching or your teaching responsibilities?

Fewer than once per month	24%
Monthly	6%
2–3 times per month	8%
Weekly	36%
More often than once per week	1%
(is not required to attend meetings)	25%

15. For about how many years in your current degree program have you been required to attend meetings such as the ones described above? Please round up (e.g., if you have attended for 0.3 years please enter 1).

N	M	SD	Min	Q1	Med	Q3	Max
164	1.8	1.1	0	1	1	2	6

16. As best as you can, please estimate the percentage of the meetings for which a faculty member or full-time lecturer is present. (e.g., if you estimate that a faculty member is present for 12 percent of the meetings, simply enter 12).

N	M	SD	Min	Q1	Med	Q3	Max
164	82%	32%	0%	80%	100%	100%	100%

17. Sometimes, even when not officially required to do so, people may voluntarily discuss topics related to teaching. The discussions may be informal and unplanned, and they may last only a short time (e.g., 30 seconds). The discussions also may occur via e-mail or other electronic methods of communication. In your current department have you ever voluntarily participated in discussions with other graduate students, lecturers, or faculty members to discuss topics related to teaching or assisting courses?

Yes	77 %
No	23%

18. On average, about how often do you participate in voluntary discussions about teaching with other graduate students, faculty members, or full-time lecturers in your department?

Fewer than once per month	17%
Monthly	14%
2–3 times per month	14%
Weekly	16%
2–3 times per week	11%
4 or more times per week	5%
(does not participate in voluntary meetings)	23%

19. For about how many years of your current degree program have you participated in voluntary discussions as described above? Please round up to the nearest year.

N	M	SD	Min	Q1	Med	Q3	Max
168	2.2	1.9	0	1	2	3	20

20. As best as you can, please estimate the percentage of voluntary discussions that include a faculty member or full-time lecturer.

N	M	SD	Min	Q1	Med	Q3	Max
168	24%	26%	0%	5%	12.5%	31%	100%

21. Please estimate the number of graduate students in your department for which you would feel comfortable approaching to start a voluntary discussion about a topic related to teaching.

N	M	SD	Min	Q1	Med	Q3	Max
218	16	18	0	5	10	20	100

22. Please estimate the number of faculty and full-time lecturers in your department for which you would feel comfortable approaching to start a voluntary discussion about a topic related to teaching.

N	M	SD	Min	Q1	Med	Q3	Max
218	7	10	0	3	5	7	90

Section 3: Faculty Support for Your Teaching*
Please consider the course, lab, or discussion section you most recently taught, assisted, or graded papers for. Indicate whether a faculty member or full-time lecturer has fulfilled each of these.

	Yes	No	N/A or chose not to respond
23. Clearly communicated my tasks, deadlines, and responsibilities for my teaching-related position.	74%	8%	18%
24. Clearly communicated to me the most important course topics to focus upon.	67%	16%	18%
25. Facilitated productive conversations with me and other graduate students regarding our teaching.	57%	25%	18%

**note these items were only asked of participants who indicated they have had some teaching responsibilities (Items 4–9).*

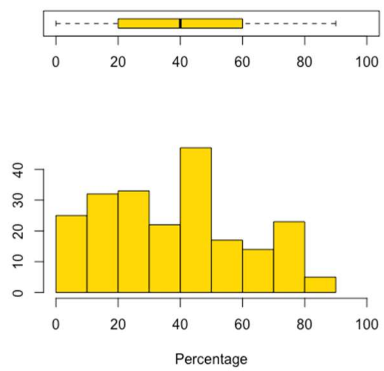
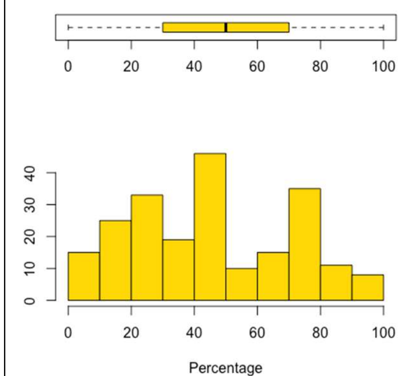
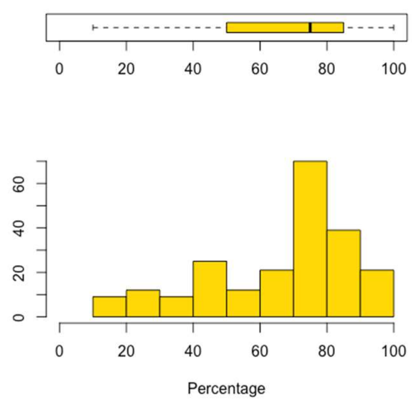
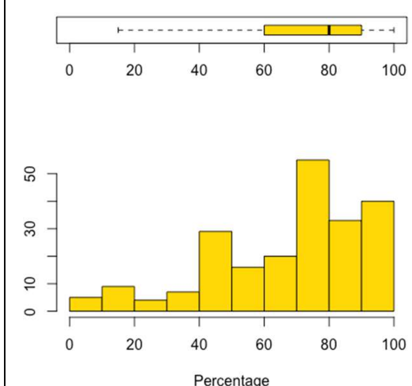
26. In your current program, how many times has a faculty member or full-time lecturer ever observed you as you were completing your teaching responsibilities and provided feedback intended to help you improve? (If you have not been observed please enter 0).

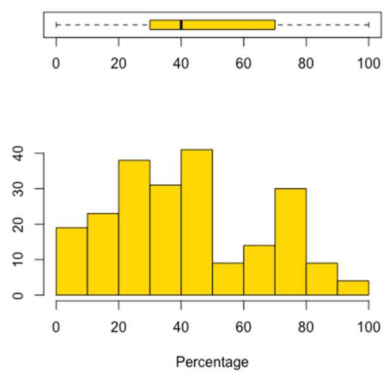
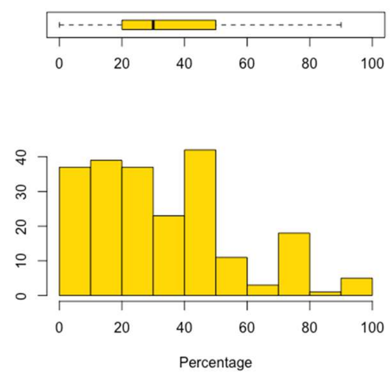
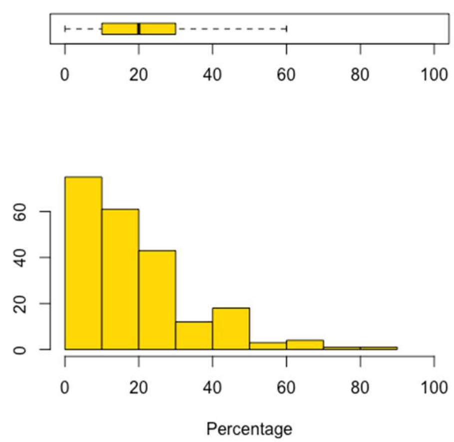
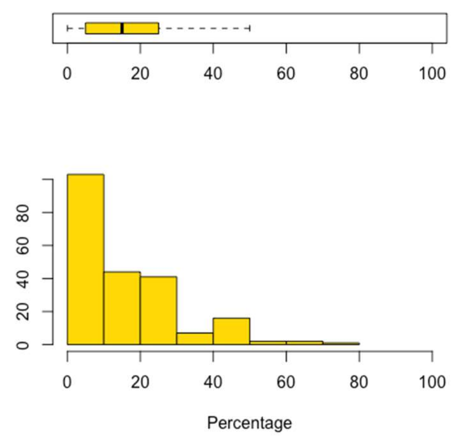
N	M	SD	Min	Q1	Med	Q3	Max
180	1.4	3.1	0	0	0	2	32

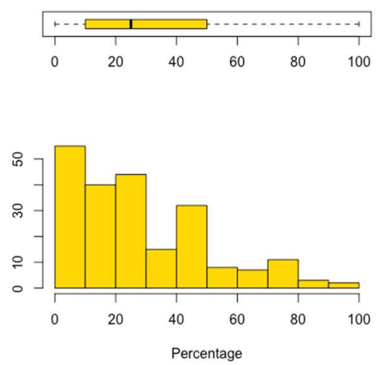
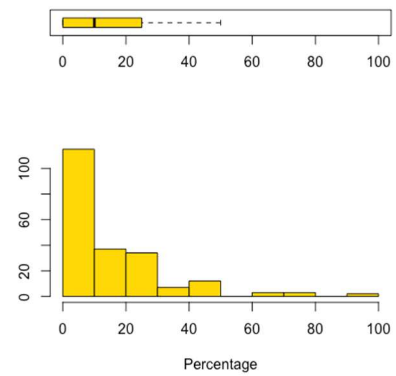
**note this item was only asked of participants who indicated they have had some teaching responsibilities (Items 4–9).*

Section 4: Your Beliefs About Teaching and Learning

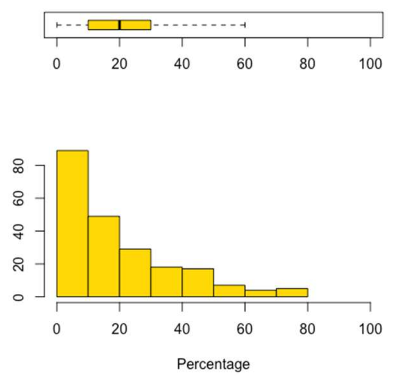
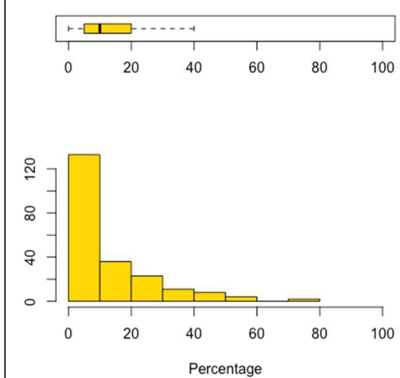
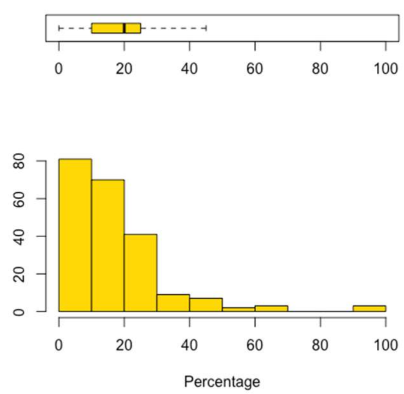
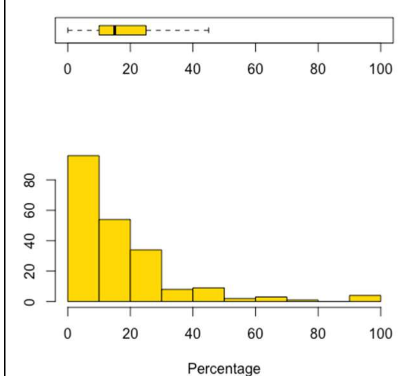
Imagine that you are asked to teach an introductory statistics course. You have the freedom to teach the course however you believe is best for student learning. There are about 35 students in the course. You are the primary, sole instructor (there is no additional lab or discussion section). The students are taking the introductory statistics course as a general university requirement, and no calculus prerequisite is required. The course is not offered online (you meet with students face-to-face). In the first column, please enter a percentage that reflects your current beliefs about how such a course should be taught. In the second column, please reflect back to the time before you entered your current degree program and enter a percentage that reflects your beliefs at that time. Please enter just the number (e.g., For 80%, please enter the number 80). The percentages do not need to add to 100. If you cannot make an estimate you may leave that space empty.

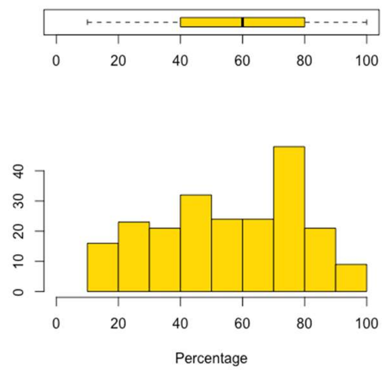
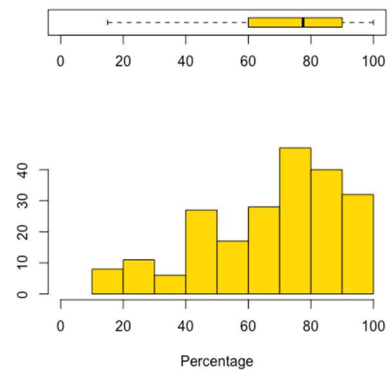
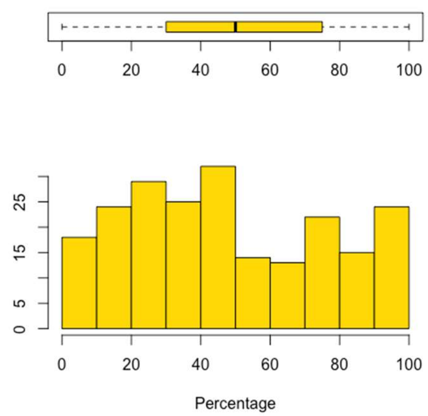
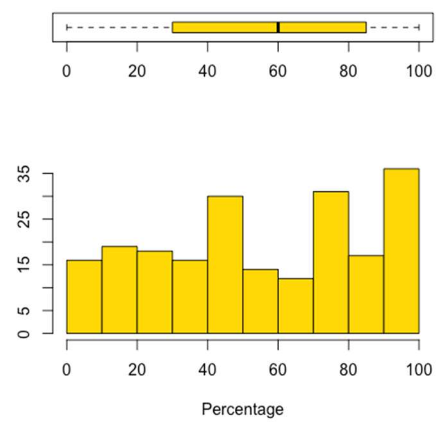
	Current Beliefs *M* (*SD*)	Beliefs Before Entering Degree Program *M* (*SD*)
27/28. What percentage of students' homework problems should be focused on practicing procedures using formulas (e.g., the formula for the standard deviation or standard error)?	42% (23%) 	50% (25%) 
29/30. For what percentage of total scores on exams and quizzes should you (as the instructor) have talked about a similar example in-class?	70% (21%) 	72% (23%) 

	Current Beliefs *M* (*SD*)	Beliefs Before Entering Degree Program *M* (*SD*)
31/32. What percentage of total scores on exams and quizzes should be based on students' explanations of their reasoning using words?	46% (24%) 	37% (23%) 
33/34. What percentage of students' course grades should be based on assessments that were completed in small groups (e.g., group quizzes, assignments, or projects)?	22% (17%) 	18% (16%) 

	Current Beliefs *M* (*SD*)	Beliefs Before Entering Degree Program *M* (*SD*)
35/36. For what percentage of the course content about statistical inference should students use simulation methods (e.g., randomization tests, bootstrapping)?	31% (23%)  	17% (19%) 

Please continue to enter numbers corresponding to percentages that reflect your beliefs about the best way to teach the introductory statistics course described on the previous page. There are about 35 students in the course. You are the primary, sole instructor (there is no additional lab or discussion section). The students are taking the introductory statistics course as a general university requirement, and no calculus prerequisite is required. The course is not offered online (you meet with students face-to-face). The percentages do not need to add to 100.

	Current Beliefs *M* (*SD*)	Beliefs Before Entering Degree Program *M* (*SD*)
37/38. What percentage of class time should be used for students to communicate their ideas to each other together in small groups?	23% (19%)  	15% (15%) 
39/40 For what percentage of class time should you (as the instructor) explain to students why their misconceptions are incorrect?	21% (16%)  	20% (18%) 

	Current Beliefs *M* (*SD*)	Beliefs Before Entering Degree Program *M* (*SD*)
41/42. What percentage of class time should be used for you (as the instructor) to present to the class (e.g. conducting demonstrations or lectures)?	59% (23%)  	71% (22%) 
43/44. For what percentage of class sessions should you (as the instructor) explain the key statistical content before students use it to solve problems?	52% (29%)  	59% (30%) 

Section 5: The People Who Influence Your Teaching Beliefs

The next ten questions ask you to consider your interactions with the two people at your institution who have had the most influence on your teaching beliefs. They may be colleagues, friends, supervisors, or mentors. They may be graduate students, faculty, or other employees at your university. Before we begin these questions, please choose the two people at your university who have the most influence on your teaching beliefs. Please designate them as Person #1 and Person #2, where Person #1 has the most influence. If you believe that nobody at your current institution has had an influence on your teaching beliefs, please choose Person #1 and Person #2 according to whom you interact with the most regarding teaching or teaching-related responsibilities.

45. Please indicate the primary role of Person #1, who has the most influence on your teaching beliefs.

Faculty or Lecturer	83%
Graduate Student	13%
Other (Please Specify) ___________________	3%

46. Please indicate the primary role of Person #2, who has the second-most influence on your teaching beliefs.

Faculty or Lecturer	56%
Graduate Student	41%
Other (Please Specify) __________________	2%
(chose not to respond)	1%

Please mark whether you agree or disagree with the following statements regarding your interactions with Person #1 from above.

	Agree	Disagree	Chose not to respond
47. I feel comfortable engaging in discussion regarding teaching topics.	95%	4%	1%
48. I feel that my point of view regarding teaching topics is acknowledged.	91%	7%	2%
49. If I disagree with Person #1, (s)he will still respect my ideas about teaching.	91%	7%	2%
50. Person #1 cares about the quality of my teaching.	90%	8%	2%
51. I admire Person #1 as a teacher.	94%	5%	1%

Please mark whether you agree or disagree with the same statements regarding your interactions with Person #2 from above.

	Agree	Disagree	Chose not to respond
52. I feel comfortable engaging in discussion regarding teaching topics.	89%	9%	2%
53. I feel that my point of view regarding teaching topics is acknowledged.	88%	10%	2%
54. If I disagree with Person #2, (s)he will still respect my ideas about teaching.	84%	14%	2%
55. Person #2 cares about the quality of my teaching.	84%	14%	2%
56. I admire Person #2 as a teacher.	83%	15%	2%

The next questions are completed for Person #1 only. Imagine that Person #1 is teaching an introductory statistics course to a class of about 35 students. The course is not calculus-based, and students are taking the course as a general university requirement (i.e., the students' majors do not have a statistics requirement). As best as you can, please indicate whether each of the following matches how you imagine that Person #1 would teach the course.

	Yes	No
57. The content is presented mostly through the instructor or TA's lectures.	72%	27%
58. This course frequently requires students to practice procedures using formulas.	72%	27%
59. The content is presented mostly through small-group activities.	38%	61%
60. Simulation methods (e.g., randomization tests, bootstrapping) are a primary tool used for teaching statistical inference.	52%	47%
61. Students have the chance to communicate their ideas using a variety of means and media (e.g., clickers, chalk boards, oral presentations).	58%	41%
62. Student's grades are calculated primarily based on in-class quizzes and exams completed individually (i.e., not completed in groups).	84%	15%

Almost Finished! Section 6: Getting to Know You

63. Prior to becoming a graduate student at your current institution, did you have experience as a full-time teacher in an elementary or secondary classroom?

Yes, for about how many years? (Please round up to the nearest year.) _____________	9%
No	91%

Summary of number of years for those who responded "Yes…"

N	M	SD	Min	Q1	Med	Q3	Max
23	2.4	2.4	0	1	1	3	10

64. Prior to becoming a graduate student at your current institution, did you have experience as a full-time instructor at a college or university?

Yes, for about how many years? (Please round up to the nearest year.) _____________	10%
No	90%

Summary of number of years for those who responded "Yes…"

N	M	SD	Min	Q1	Med	Q3	Max
21	3.95	4.75	0	2	2	4	22

65. Have you ever acquired a student visa in order to attend a college or university in the United States?

Yes	25 %
No	75%

66. In which institution are you enrolled? Please type the full name and do not use abbreviations.

(See Appendix K: Institutions Represented)

67. Do you intend to earn a Ph.D. at this institution?

Yes	77 %
No	16%
Undecided	7%

68. Do you expect to teach statistics courses as a part of your primary career?

Yes	47 %
No	53%

69. What is your age?

N	M	SD	Min	Q1	Med	Q3	Max
216	26.8	5	21	24	26	28	54

70. With which gender do you most identify yourself?

Male	51 %
Female	49 %
I do not identify as male or female. I identify as: _________________	<1 %

(71). In the space below, please offer any additional comments or information you wish to provide.

Entry into Prize Drawing

Thank you so much for your participation! You are now eligible to participate in a random drawing for a $25 amazon.com gift card. Five winners will be selected. If you would like to be entered, please enter your name and e-mail address below. Data will be de-identified before being analyzed and published.

	Your Contact Information
Please enter your name	
Please enter your e-mail address	

Please click the NEXT button to submit your entry and complete the survey.

APPENDIX C: Focus Group Questions

1. Topic: Discussions about teaching

Prompt: Do you tend to talk to other GTAs in your program about teaching? Think of 1–3 graduate teaching assistants (GTAs) with whom you have talked about your teaching over the past term.

Questions:

a. What usually leads to these discussions? What kinds of things do you talk about?
b. How often do the discussion happen?
c. How do the topics vary over the term?
d. Is it a set group of GTAs or do other GTAs to join your discussions? How do they initiate joining?
e. What are the characteristics of the GTAs you tend to talk to the most about teaching? Are they the same age/year in program? Do they teach the same class? Is their desk near yours? Are you friends outside of the program?
f. Do you feel like you are in a group? If so, what defines the group? Who is able to be part of it, who does not tend to join in?
g. Where do your discussions about teaching usually take place?
h. Do the discussions typically happen during regular meetings or outside of regular meetings or both?

2. Topic: GTAs seeking help from other GTAs

Prompt: Have you had any challenges, difficulties, or surprises regarding teaching? Think of 1–3 challenges you have faced as a teacher of statistics and think about what you did to get help or advice about these challenges.

Questions:

a. If so, who did you turn to for help or advice?
b. Did you turn to other GTAs for solutions?
c. Is there anyone else you thought about going to for help/advice?
d. How helpful were the people you sought out for advice or help?
e. Are there certain characteristics of GTAs that make them more likely for you to discuss teaching topics and issues with?
f. Do you feel comfortable discussing issues related to teaching with other GTAs in your department outside of structured meetings?

3. Topic: GTAs giving help to other GTAs

Prompt: Have any other GTAs approached you with challenges or questions regarding teaching statistics? Try to think of 1–3 occasions when you have been approached by another GTA.

a. What were the settings in which they approached you? (Was it in a shared office, in a class, outside of class, on the phone, etc.?)
b. What was the challenge or problem they were dealing with?
c. Were you able to help the GTA or provide a solution? How did you do this?
d. Were other GTAs involved in discussing this issue?

4. Topic: Faculty and Department Support for GTAs.

Prompt: How do the faculty support your teaching of statistics? Think of 2–3 occasions when you have discussed teaching statistics with faculty members.

   a. What were the settings in which the discussions occurred? (Was it in a shared office, in a class, outside of class, on the phone, etc.?)
   b. Do the discussions typically happen during regular meetings or outside of regular meetings or both?
   c. What types of topics related to teaching statistics have you discussed with faculty members?
   d. Are there certain characteristics of faculty that make them more likely for you to discuss teaching topics and issues with?
   e. How would you describe the way that the GTAs in your department interact with faculty regarding teaching?
   f. What could your department do to allow or support GTAs to support one another more regarding teaching?

5. Topic: The culture of GTA support for one another.

Prompt: What is the overall culture of the GTAs in your department regarding teaching? Think about how you would communicate the atmosphere in your department regarding teaching, and the extent to which student-learning is the goal of interactions regarding teaching statistics.

Questions:

   a. What would you say are the different goals of GTAs in your department regarding teaching?
   b. How would you describe the overall atmosphere of GTAs in your department regarding teaching?
   c. Do you have any other comments or questions you would like to share?

APPENDIX D: Survey Blueprint

Construct	No. of Items	Drafted Item(s)	Response Type	Item Source (if applicable)
1. Perceived Faculty Presence in CoP	4	Please rate the extent to which you agree or disagree that a faculty member or full time lecturer has fulfilled each of these. a. Clearly communicated my tasks, deadlines, and responsibilities for my teaching-related position. b. Clearly communicated to me important topics or goals for my teaching-related position. c. Helped facilitate productive dialogue with me and other graduate students regarding our teaching.	Likert	Arbaugh et al. (2008).
		In your current program, has a faculty member ever observed your teaching and provided feedback intended to help you improve your teaching?	Yes/No	N/A
2. Perceived CoP Norms of Interaction	4	Please consider the faculty and graduate students you have interacted with most often regarding your teaching responsibilities in your current degree program. Rate the extent to which you agree or disagree with these statements regarding these interactions. a. I feel comfortable engaging in discussion with others regarding teaching topics. b. I feel that my point of view regarding teaching topics is acknowledged by others. c. I feel comfortable disagreeing with others regarding teaching topics while still maintaining a sense of respect. d. I feel that others care about me in terms of the quality of my teaching and the success of my students.	Likert	Arbaugh et al. (2008)

Construct	No. of Items	Drafted Item(s)	Response Type	Item Source (if applicable)
3. Perceived CoP Orientation toward Student-Centered Teaching	4	Consider a student who is fully engaged in an introductory statistics course offered by the person in your department (graduate student or faculty member) who has the most influence on your teaching. Indicate the extent to which you think that student would agree or disagree with each of the following statements. a. The content was presented mostly through the instructor or TA's lectures. b. The course frequently required students to work together. c. This course required students to do a lot of practice of procedures using formulas. d. Quizzes and exams were used as the primary way to evaluate student learning.	Likert	Justice et al. (in press); Zieffler et al. (2012)
4. Perceived Engagement in CoP	4	1. Do you currently have a common or shared space (e.g., office or cubicle) with other graduate students who are involved with the teaching or assisting of a statistics course?	Yes/No	N/A
		2. Do you feel a spirit of community with at least one other graduate student who has teaching responsibilities in my department?	Yes/No	Rovai (2002)
		3. About how long, if at all, would you consider yourself to have been a part of a community of 2 or more people who meet somewhat regularly to discuss topics regarding teaching?	Numerical	N/A
		4. Typically, about how many days per week do you discuss teaching-related topics with other graduate students, either in formal meetings or informal conversations?	Numerical	N/A
5. Current Teaching Practices	4	Consider a student who is fully engaged in your introductory statistics course. Indicate the extent to which you think that student would agree or disagree with each of the following statements. a. The content is presented mostly through the instructor or TA's lectures. b. The course frequently requires students to work together. c. This course requires students to do a lot of practice of procedures using formulas. d. Quizzes and exams are used as the primary way to evaluate student learning.	Likert	Justice et al. (in press); Zieffler et al. (2012)

Construct	No. of Items	Drafted Item(s)	Response Type	Item Source (if applicable)
6. Current Beliefs about Teaching Statistics	4	Please rate the extent to which you agree or disagree with each of the following statements as they reflect your beliefs (but not necessarily your actual teaching) of an introductory course for students (non-majors) taking statistics as a general requirement. a. Lectures should be the primary way for students to learn statistical content. b. Students should frequently be required to work together in small groups. c. Mathematical formulas (e.g., the formula for calculating the standard deviation or the standard error) should play a primary role. d. Quizzes and exams should be used as the primary way to evaluate student learning.	Likert	Justice et al. (in press); Zieffler et al. (2012)
7. Beliefs about Teaching Statistics, before entering CoP	4	Please reflect back to the time just **before** entering your current graduate program. Rate the extent to which you think you would have agreed or disagreed with each of the following statements as they reflected your beliefs (but not necessarily your actual teaching) of an introductory course for students (non-majors) taking statistics as a general requirement. a. Lectures should be the primary way for students to learn statistical content. b. Students should frequently be required to work together in small groups. c. Mathematical formulas (e.g., the formula for calculating the standard deviation or the standard error) should play a primary role. d. Quizzes and exams should be used as the primary way to evaluate student learning.	Likert	Justice et al. (in press); Zieffler et al. (2012)
8. International Student Status	1	Do you consider yourself to be an international student?	Yes / No	GSSTI 2
9. Interest in Teaching	1	Do you intend to pursue a career that largely involves teaching?	Forced-choice	GSSTI 2
10. Previous Teaching Experience	2	Prior to becoming a graduate student at your current institution, did you have experience as a full-time instructor in an elementary or secondary classroom?	Yes / No	N/A
		Have you ever been a full-time instructor at a college or university?	Yes / No	N/A

Construct	No. of Items	Drafted Item(s)	Response Type	Item Source (if applicable)
11. Year in Program	1	Please indicate the number of years you have completed in your current graduate degree program. Round to the nearest half-year (e.g., .25 rounds to .5).	Numerical	N/A
12. Target population check	9	1. Are you currently enrolled as graduate student?	Yes / No	N/A
		2. In which institution are you enrolled?	Open-Response	GSSTI 2
		3. What best describes the type of department are you enrolled in?	Forced-Choice	N/A
		4. At your current institution have you ever been hired in positions that include the following teaching-related activities? a. Grading papers b. Holding office hours c. Facilitator of a weekly lab or discussion section d. Primary instructor for a course e. Assistant to a primary instructor for a course f. Overseer of other graduate students for teaching responsibilities	Yes / No	GSSTI 2

APPENDIX E: Examples of Original and Revised Items

Construct: Prior Beliefs. Original item is from an instrument used by Justice et al. (in press).

Original item:
Please rate the extent to which you agree or disagree with each of the following statements as they reflect your beliefs (but not necessarily your actual teaching) of an introductory statistics course.
Lectures should be the primary way for students to learn statistical content.

Response options: Strongly Disagree, Disagree, Agree, Strongly Agree, Undecided

Adapted item:
Imagine that you are asked to teach an introductory statistics course. You have the freedom to teach the course however you believe is best for student learning. There are about 35 students in the course. You are the primary, sole instructor (there is no additional lab or discussion section). The students are taking the introductory statistics course as a general university requirement, and no calculus prerequisite is required. The course is not offered online (you meet with students face-to-face). In the first column, please enter a percentage that reflects your current beliefs about how such a course should be taught. In the second column, please reflect back to the time before you entered your current degree program and enter a percentage that reflects your beliefs at that time. Please enter just the number (e.g., For 80%, please enter the number 80). The percentages do not need to add to 100. If you cannot make an estimate you may leave that space empty.

What percentage of class time should be used for you (as the instructor) to present to the class (e.g. conducting demonstrations or lectures)?

Construct: Norms. The original item had a factor-loading of .620 on the "Social Presence" factor in a study by Arbaugh et al. (2008).

```
Original item:
I felt comfortable disagreeing with other course participants
while still maintaining a sense of trust. (Arbaugh, et al.,
2008)

Adapted item:
If I disagree with Person #(XX), (s)he will still respect my
ideas about teaching.
```

Construct: Orientation. Original item is from an instrument used by Justice et al. (in press).

Original item:
Consider a student who is fully engaged in your introductory statistics course. Indicate the extent to which you think that student would agree or disagree with each of the following statements.
The content was presented mostly through the instructor or TA's lectures.

Adapted item:
The next questions are completed for Person #1 only. Imagine that Person #1 is teaching an introductory statistics course to a class of about 35 students. The course is not calculus-based, and students are taking the course as a general university requirement (i.e., the students' majors do not have a statistics requirement). As best as you can, please indicate whether each of the following matches how you imagine that Person #1 would teach the course.
The content was presented mostly through the instructor or TA's lectures.

Construct: Faculty Presence. The original item had a factor-loading of .633 on the "Leadership" factor in a study by Arbaugh et al. (2008).

Original item:
The instructor helped keep course participants engaged and participating in productive dialogue. (Arbaugh et al., 2008) (Likert-type response options)

Adapted item:
Please consider the course, lab, or discussion section you most recently taught, assisted, or graded papers for. Indicate whether a faculty member or full-time lecturer has fulfilled each of these.

Facilitated productive conversations with me and other graduate students regarding our teaching.

APPENDIX F: Solicitation for Participation Posted in the Caucus of Academic Representatives Weekly Digest

Request for help gathering data from your statistics graduate students

I am writing to ask if you would be willing to send your statistics graduate students an e-mail invitation (see below) to participate in an online survey that is part of my dissertation research. The survey, which should only take students about 15–20 minutes to complete, will be used to posit a model for how statistics graduate students develop as teachers.

Data collected from this study will hopefully inform the community about statistics students' experiences in their graduate programs, and may lead to the design of future professional development resources and opportunities.

If you have any questions, you are welcome and encouraged to contact me, the principal investigator, Nicola Justice (njustice@umn.edu). You may also contact my research adviser, Joan Garfield (jbg@umn.edu).

Thanks so much for your help!
Nicola Justice
Ph.D. Candidate
Statistics Education, University of Minnesota

E-mail to be sent to statistics graduate students:

Subject Line: Opportunity to participate in survey-study on statistics graduate students and win a $25 Amazon.com gift card.

I am writing to ask if you would be willing to participate in my dissertation research by completing a short survey. The survey, which should only take about 15–20 minutes to complete, will be used to posit a model for how statistics graduate students develop as teachers.

At the end of the survey, you will have the opportunity to enter to win one of five $25 Amazon.com gift cards that will be randomly awarded to participants. Moreover, your participation will contribute to research about statistics graduate students' experiences in their programs, and may be able to inform the design of future professional development resources and opportunities.

If you have any questions, you are welcome and encouraged to contact the principal investigator, Nicola Justice (njustice@umn.edu). You may also contact my research adviser, Joan Garfield (jbg@umn.edu).

Thank you for your consideration! If you would like to participate, please click the following link:

https://umn.qualtrics.com/SE/?SID=SV_bqmkQSWsSMpABg1

APPENDIX G: First E-mail Solicitation for Participation sent to CAR

From: Wasserstein, Ronald L.
Sent: Wednesday, February 17, 2016 6:36 PM
To: Caucus of Academic Representatives <STATACADREPS@amstat.org>
Subject: Request for help gathering data from your statistics graduate students

Dear Caucus of Academic Reps,
This message is sent on behalf of a statistics education Ph.D. candidate at the University of Minnesota. For her dissertation, research, she is collecting data to help posit a model for how statistics graduate students develop as teachers. She would like to survey your graduate students.

The survey will take students about 15–20 minutes to complete, and the data collected from this study will hopefully inform the community about statistics students' experiences in their graduate programs, and may lead to the design of future professional development resources and opportunities.

Please forward the message below to your grad students.
Thank you.
Ron

E-mail to be sent to statistics graduate students:

Subject Line: Opportunity to participate in survey-study on statistics graduate students and win a $25 Amazon.com gift card.

I am writing to ask if you would be willing to participate in my dissertation research by completing a short survey. The survey, which should only take about 15–20 minutes to complete, will be used to posit a model for how statistics graduate students develop as teachers.

At the end of the survey, **you will have the opportunity to enter to win one of five $25 Amazon.com gift cards that will be randomly awarded to participants**. Moreover, your participation will contribute to research about statistics graduate students' experiences in their programs, and may be able to inform the design of future professional development resources and opportunities.

If you have any questions, you are welcome and encouraged to contact the principal investigator, Nicola Justice (parke675@umn.edu). You may also contact my research adviser, Joan Garfield (jbg@umn.edu).

Thank you for your consideration! If you would like to participate, please click the following link:

https://umn.qualtrics.com/SE/?SID=SV_bqmkQSWsSMpABg1

Nicola Justice
Ph.D. Candidate

Ronald L. Wasserstein
Executive Director

American Statistical Association
Promoting the Practice and Profession of Statistics®

APPENDIX H: Second E-mail Solicitation for Participation Sent to CAR

From: Wasserstein, Ronald L.
Sent: Wednesday, February 24, 2016 6:27 PM
To: Caucus of Academic Representatives <STATACADREPS@amstat.org>
Subject: 2nd and final request for help gathering data from your statistics graduate students

Dear Caucus of Academic Reps members,

I am writing once more to ask if you would be willing to forward your statistics graduate students a second and final e-mail invitation (see below, bottom message) to participate in a **10–15 minute** online survey that is part of my dissertation research.

I am looking for about 50 more participants in order to complete my dissertation.

If you were unable to send the first e-mail invitation, I have included it below as an alternate option.

Thanks so much for your help! If you have any questions, you are welcome and encouraged to contact me, the principal investigator, Nicola Justice (njustice@umn.edu). You may also contact my research adviser, Joan Garfield (jbg@umn.edu).

Nicola Justice
Ph.D. Candidate
Statistics Education
University of Minnesota

1st e-mail to be sent to statistics graduate students (if previous e-mail was not sent last week):

Subject Line: Opportunity to participate in survey-study on statistics graduate students and win a $25 Amazon.com gift card.

I am writing to ask if you would be willing to participate in my dissertation research by completing a short survey, which typically takes **about 10–15 minutes** to complete. The data will be used to posit a model for how statistics graduate students develop as teachers.

I am currently looking for about 50 more participants in order to finish my dissertation.

To thank you for your time, **participants will have the opportunity to enter to win one**

of five $25 Amazon.com gift cards that will be randomly awarded after the survey closes. Moreover, your participation will contribute to research about statistics graduate students' experiences in their programs, and may be able to inform the design of future professional development resources and opportunities.

If you have any questions, you are welcome and encouraged to contact the principal investigator, Nicola Justice (njustice@umn.edu). You may also contact my research adviser, Joan Garfield (jbg@umn.edu).

Thank you for your consideration! If you would like to participate, please click the following link:

https://umn.qualtrics.com/SE/?SID=SV_bqmkQSWsSMpABg1

Nicola Justice
Ph.D. Candidate
Statistics Education
University of Minnesota

2nd and Final E-mail to be sent to statistics graduate students (if first e-mail was sent):

Subject Line: Final Call: Opportunity to win a $25 Amazon.com gift card and help a Ph.D. student gather dissertation data

I am writing in hopes of soliciting 50 more statistics graduate students who would be willing to complete survey as part of my dissertation research. The survey typically takes students 10–15 minutes to complete.

If you have already participated in the survey – thank you! You do not need to take the survey again.

If you have not yet had a chance to take the survey, please consider participating. To thank you for your time, **participants will have the opportunity to enter to win one of five $25 Amazon.com gift cards** that will be randomly awarded after the survey closes.

Thank you for your consideration! If you would like to participate, please click the following link:

https://umn.qualtrics.com/SE/?SID=SV_bqmkQSWsSMpABg1

If you have any questions, you are welcome and encouraged to contact the principal investigator, Nicola Justice (njustice@umn.edu). You may also contact my research adviser, Joan Garfield (jbg@umn.edu).

Nicola Justice
Ph.D. Candidate
Statistics Education
University of Minnesota

Ronald L. Wasserstein
Executive Director

American Statistical Association
Promoting the Practice and Profession of Statistics®

APPENDIX I: Example of First E-mail Solicitation to Faculty Contacts

Dear _______,

Joan Garfield (my Ph.D. adviser) suggested that because of your support for the DEFT project proposal we submitted (but was not funded), you might also be willing to help me collect data for my dissertation research. The goal of my research is to create and validate a model that describes how statistics graduate students develop as teachers.

I am writing in hopes that you would be willing to forward the message below to your statistics graduate students, inviting them to participate in an online survey that typically takes about 10–15 minutes to complete. To thank them for their time, participants will have the opportunity to enter their names in a random drawing for one of five $25 Amazon.com gift cards.

If you have any questions, you are welcome to contact me, Nicola Justice (njustice@umn.edu). You may also contact my research adviser, Joan Garfield (jbg@umn.edu).

Thanks so much for your help!

Nicola Justice
Ph.D. Candidate
Statistics Education
University of Minnesota

E-mail to be sent to statistics graduate students:

Subject Line: Opportunity to participate in survey-study on statistics graduate students and win a $25 Amazon.com gift card.

I am writing to ask if you would be willing help me complete my dissertation research by participating in an online survey. The survey, which typically takes about 10–15 minutes to complete, will be used to posit a model for how statistics graduate students develop as teachers.

To thank you for your time, at the end of the survey, **you will have the opportunity to enter to random drawing to win one of five $25 Amazon.com gift cards**. Moreover, your participation will contribute to research about statistics graduate students' experiences in their programs, and may be able to inform the design of future professional development opportunities.

If you have any questions, you are welcome to contact me, Nicola Justice (njustice@umn.edu). You may also contact my research adviser, Joan Garfield (jbg@umn.edu).

Thank you for your consideration! If you would like to participate, please click the following link:

https://umn.qualtrics.com/SE/?SID=SV_bqmkQSWsSMpABg1

Nicola Justice
Ph.D. Candidate
Statistics Education
University of Minnesota

APPENDIX J: Example of Follow-up E-mail Solicitation to Faculty Contacts

Dear ________ ,

I hope you are doing great!

I wonder whether you were able to pass on to your statistics graduate students the invitation to participate in my dissertation research. I don't yet see any participants from your institution, and I would really love for them to be represented in the sample.

If you are willing, below I have included a follow-up e-mail that could be sent to statistics graduate students in your department. As of today, I am in need of only about 20 more participants!

Either way, thanks so much for your time!

Sincerely,
Nicola Justice
Ph.D. Candidate
Statistics Education
University of Minnesota

____

Follow-up E-mail

Subject Line: Final Call: Opportunity to win a $25 Amazon.com gift card and help Ph.D. student gather dissertation data

I am writing in hopes of soliciting 20 more statistics graduate students who would be willing to complete survey and help me complete my dissertation research. The survey typically takes students 10–15 minutes to complete.

To thank you for your time, **participants will have the opportunity to enter to win one of five $25 Amazon.com gift cards** that will be randomly awarded after the survey closes.

Thank you for your consideration! If you would like to participate, please click the following link:

https://umn.qualtrics.com/SE/?SID=SV_bqmkQSWsSMpABg1

If you have any questions, you are welcome and encouraged to contact the principal investigator, Nicola Justice (njustice@umn.edu). You may also contact my research adviser, Joan Garfield (jbg@umn.edu).

APPENDIX K: Institutions Represented in Sample Used to Compute Final Model

Index	Institution	US State	Region
1	Baylor University	Texas	Southwest
2	Brown University	Rhode Island	Northeast
3	Carnegie Mellon University	Pennsylvania	Northeast
4	Columbia University	New York	Northeast
5	Duke University	North Carolina	Southeast
6	Emory University	Georgia	Southeast
7	Iowa State University	Iowa	Midwest
8	Johns Hopkins University	Maryland	Northeast
9	Medical University of South Carolina	South Carolina	Southeast
10	Montana State University	Montana	West
11	North Carolina State University	North Carolina	Southeast
12	Ohio State University	Ohio	Midwest
13	Penn State University	Pennsylvania	Northeast
14	Purdue University	Indiana	Midwest
15	Rice University	Texas	Southwest
16	State University of New York, Buffalo	New York	Northeast
17	Texas A&M University	Texas	Southwest
18	Truman State University	Missouri	Midwest
19	University of California, Berkeley	California	West
20	University of California, Irvine	California	West
21	University of California, Los Angeles	California	West

Index	Institution	US State	Region
22	University of Georgia	Georgia	Southeast
23	University of Iowa	Iowa	Midwest
24	University of Kentucky	Kentucky	Southeast
25	University of Massachusetts, Amherst	Massachusetts	Northeast
26	University of Michigan, Ann Arbor	Michigan	Midwest
27	University of Minnesota	Minnesota	Midwest
28	University of Nebraska, Lincoln	Nebraska	Midwest
29	University of New Mexico	New Mexico	Southwest
30	University of North Carolina, Chapel Hill	North Carolina	Southeast
31	University of Rochester	New York	Northeast
32	University of South Carolina	South Carolina	Southeast
33	University of Texas, Austin	Texas	Southwest
34	University of Utah	Utah	West
35	University of Washington	Washington	West
36	University of Wisconsin, Madison	Wisconsin	Midwest
37	Vanderbilt University	Tennessee	Southeast

Note. N=212. Some participants included in the final model did not indicate their institution (n=6).

APPENDIX L: Correlation Matrix of Scores for the Six Core Constructs

	Measure	1	2	3	4	5	6
1.	Current Beliefs	—					
2.	Prior Beliefs	.48	—				
3.	Engagement	.16	.03	—			
4.	Norms of Interaction	–.04	–.11	.00	—		
5.	Orientation	.05	.00	.02	.21	—	
6.	Faculty Presence	.12	.04	.66	.19	.07	—

Note. Pairwise-complete Pearson correlations are given. $N = 218$ unless with Norms of Interaction ($N = 216$) and Orientation ($N = 217$).

APPENDIX M: R Code for Cross-Validation and Model Selection

```
library(stats)
rm(list = ls(all = TRUE))

df.model.na.removal<-function(mod,df){
 #input is a data frame and the model
 #output is the data frame with the NA's removed for that model
 df1<-df
 df1$fitted<-fitted(lm(mod,data=df0,na.action=na.exclude))
 dfNAs<-df1[(is.na(df1$fitted)),]
 #df2 is the remaining data after NAs from model are removed
 #df2 is the set that will be used for cross validatation
 #df2 will change for each model used.
 df2<-dfremaining<-df1[!is.na(df1$fitted),]
 stopifnot(dim(dfNAs)[1]+dim(dfremaining)[1]==dim(df0)[1])
 return(df2)
}

df.rearrange<-function(df){
 #input is a data frame
 #output is a data frame reordered randomly
 sample.vector<-sample(1:dim(df)[1],dim(df)[1],replace=F) #creates a vector of random
new indices
 df$sample.vector<-sample.vector
 df1<-df[order(sample.vector),]
 return(df1)
}

fold.length.vector<-function(fold,df){
 #input is a data frame and the number of folds
 #output is a vector with the size of each fold
 #this function accounts for the fact that the remainder will not always be 0.
 vec<-rep(0,fold)
 N<-dim(df)[1]
 remainder<-N%%fold
 for (j in 1:fold){
 vec[j]<-ifelse(N %% fold==0,
            N/fold,
            ifelse(j>remainder,
                floor(N/fold),
                floor(N/fold)+1))
 }
```

```
  return(as.vector(vec))
}

fold.starts.vector<-function(fold,df){
  #input is the # of folds, and the data frame.
  #output is a vector that gives the indices of the start of each fold
  vec<-rep(0,fold)
  length.vec<-fold.length.vector(fold,df)
  for (i in 1:fold){
    vec[i]<-ifelse(i>1,
                1+sum(length.vec[1:i-1]),
                1)
  }
  return(vec)
}

fold.ends.vector<-function(fold,df){
  #input is the # of folds, data frame
  #output is the end indices for each fold
  vec<-rep(0,fold)
  length.vec<-fold.length.vector(fold,df)
  for (i in 1:fold){
    vec[i]<-sum(length.vec[1:i])
  }
  return(vec)
}

cv.once<-function(mod,train.set,test.set){
  #performs the cross-validation procedure for one respliting
  #input is a model, the training set, the test set
  #Note: sets should already be chosen and NAs should be removed.
  #output is c(MSE,NPJ.AIC,NPJ.AICc,NPJ.BIC)
  #MSE is sum of squared residuals divided by n
  trained.model<-lm(mod,data=train.set);
  #apply trained model to test set
  test.set.y.hat<-predict(trained.model,newdata=test.set,type="response")
  #calculate RSS
  n<-length (test.set.y.hat)
  RSS2<-sum((test.set$bel-test.set.y.hat)^2)
  MSE<-RSS2/n
  #calculate model selection criteria
  ncoefs<-length(lm(mod,data=train.set,na.action=na.exclude)$coef)
  NPJ.AIC<-n*log(RSS2/n,exp(1))+2*ncoefs
  NPJ.AICc<-n*log(RSS2/n,exp(1))+2*ncoefs+
```

```
   (2*ncoefs*(ncoefs+1))/(n-ncoefs-1)
  NPJ.BIC<-n*log(RSS2/n,exp(1))+log(n,exp(1))*(ncoefs)
  return(c(MSE,NPJ.AIC,NPJ.AICc,NPJ.BIC))
}

#create data to put in the CV once function to test it out
#first define a model and # of folds
model<-'bel~pbel*ori'
fold<-5
#df0 was my data frame for testing.

#try out some of the functions.
dim(df0)
df1<-df.model.na.removal(mod=model,df=df0);dim(df1)
df2<-df.rearrange(df1);dim(df2)
flv<-fold.length.vector(fold,df2)
fsv<-fold.starts.vector(fold,df2)
fev<-fold.ends.vector(fold,df2)
te.s<-df2[fsv[3]:fev[3],];dim(te.s) #I just chose 4th fold for test set
tr.s<-df2[-(fsv[3]:fev[3]),];dim(tr.s)

#try out cv.once function
cv.once(model,tr.s,te.s)
#Everything looks good

cv.across.folds.once<-function(mod,fold,df){
  #this function does the cross-validation procedure for all the folds in one resplitting
  #input is a model, # of folds, and data frame
  #data frame should already have NAs removed
  # this chooses the training set and test set for each of the folds
  #output is a data frame with each row representing the resuts from one fold
  # and columns (with the results) are:
  #MSE for the test set using the train set coefficients for that fold
  #AIC, AICc, and BICc calculated for applying the train set coefficients to the test set.
  length.vec<-fold.length.vector(fold,df)#;length.vec
  starts.vec<-fold.starts.vector(fold,df)#;starts.vec
  ends.vec<-fold.ends.vector(fold,df)#;ends.vec
  out.df<-as.data.frame(matrix(0,ncol=4,nrow=fold))#;out.df
  names(out.df)<-c("MSE","AIC","AICc","BIC")
  for (i in 1:fold){#think of i as the fold number within the k folds
    test<-df[starts.vec[i]:ends.vec[i],];head(test);dim(test)#defines the df that makes up the
test set
    train<-df[-(starts.vec[i]:ends.vec[i]),];head(train);dim(train)#defines the df that makes
up the train set
```

```
    train$phd
    test$phd
    predict(lm(mod,data=train),newdata=test,type="response")
    cv.for.i<-cv.once(mod,test,train);cv.for.i
    out.df[i,1]<-cv.for.i[1] #prints the RSS in column 1, row i
    out.df[i,2]<-cv.for.i[2] #prints the AIC in column 1, row i
    out.df[i,3]<-cv.for.i[3] #prints the AICc in column 1, row i
    out.df[i,4]<-cv.for.i[4] # prints the BIC in column 1, row i
  }
  return(out.df)
}

cv1<-cv.across.folds.once(model,5,df2);cv1
#cv1 should be a matrix of length 5, if fold was set to 5.

cv.iterations<-function(mod,fold,iterations,df){
  #this does the cross-validation procedure for many resplittings
  #returns a data frame where each column is a type of result (MSE, AIC, AICc or BIC)
  #each row represents results from a test set on trained data.
  #there will be iterationsXfolds rows because they're all rbinded together.
  set.seed(16)
  df.nas.removed<-df.model.na.removal(mod,df)
  #seeds.vector<-sample(iterations*10,size=iterations,replace=FALSE) #samples from
1:10*iterations; sets seed for each iteration
  #set.seed(seeds.vector[1])
  out<-cv.across.folds.once(mod,fold,df.nas.removed)
  for(i in 2:iterations){
    df1<-df.rearrange(df.nas.removed)
    #set.seed(seeds.vector[i])
    add.out<-cv.across.folds.once(mod,fold,df1)
    out<-rbind(out,add.out)
  }
return(out)
}

#choose a number of iterations
iterations<-100

#test out the iterations function
cv.iterations(model,fold,iterations,df0)
#should give a matrix of length fold*iterations

cv<-function(mod,fold,iterations,df){
  #the whole cross validation process is done by cv iterations
```

```
  # this function just calculates the summaries and gives them instead of the big long df of
the 4 columns.
  input<-cv.iterations(mod,fold,iterations,df)
  means<-apply(input,2,mean,na.rm=FALSE) # gives four means.  One for all the sigma
squareds, one for all the AICs, one for all the AICcs, one for all the BICs...
  stdevs<-apply(input,2,sd,na.rm=FALSE) #analagous to means.  Gives Standard
deviations for the four measures.
  out.df<-rbind(means,stdevs) #output data frame
  return(out.df)
}




}
```