The Relationship Between Faculty Teaching Preparation and Student Ratings of Teaching

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Margaret Fitzgerald-Sisk

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

Adviser: James Brown

October, 2011

Acknowledgements

I want to acknowledge and thank the individual members of my committee. Brad Greiman, committee member and supporter, was available at the oddest of times to listen to some not-so-interesting ideas I had at the beginning, and he helped me distill those ideas into something much more workable. Ilene Alexander, committee member and SRT SME extraordinaire, helped me jump through some hoops, listened to me talk through too many silly (and off-track) ideas, and helped me speculate on the causes of some surprising findings. Judy Lambrecht, professor emerita, my defense committee chair, agreed to stay with me on this project even though she retired. She was instrumental in making this more than I expected it could be. Finally, Jim Brown, my adviser, was assigned to me from the start, and I have been so thankful. We found, over the years, that we had more in common than we expected, and he was instrumental in helping me understand how the system works. For that, he will have my eternal gratitude.

Dedication

This dissertation is dedicated to my husband, David, who, 18 years ago, promised me that if I supported him while he finished his doctoral degree, he would repay me tenfold. He has.

Table of Contents

List of Tables vi

List of Figures viii

Chapter One: Introduction 1

Graduate Programs in the United States 2

Literature on Graduate Education 3

The Job of the Professor 5

Preparing Future Faculty Programs 6

Researchable Problem Statement 8

Need 8

Hypotheses 9

Conclusion 9

Chapter Two: Literature Review 11

How Faculty Learn to Teach 11

What Are Teaching and Learning? 12

Research vs. Teaching 17

An Exploration of the Professoriate in Europe and the UK 18

The Professoriate in Europe 19

European Universities and Graduate Degree Programs 23

Quality Assurance and the European Professoriate 24

The Case of the Czech Republic 25

German Universities and Doctoral Education 26

Quality Assurance in Higher Education 28
The Spellings Commission 30
Reactions to the State of Teaching in Higher Education 31
Teaching Effectiveness 34
Conceptual Framework 35
Conclusion 40
Chapter Three: Research Design and Methodology 43
Research Plan 43
Methodology and Methods 44
Target Population 45
Sampling Procedure 46
Data Analysis Procedures 49
Ethics and Human Relations 53
Strengths and Weaknesses 54
Chapter Four: Results of the Study 56
SRT Question One 60
SRT Question Two 63
SRT Question Three 65
SRT Question Four 67
Chapter Five: Discussion, Conclusions, and Implications 70
Regression Equations 70
Skewedness 73

Sample Size 73

Database Skewedness 73

Sample Skewedness 77

Further Questions 78

Future Research 79

References 82

Appendix A: Survey Instrument 95

Appendix B: General Results Tables 101

Appendix C: SRT Question One Results Tables 104

Casewise Diagnostics 105

Appendix D: SRT Question Two Results Tables 110

Casewise Diagnostics 112

Appendix E: SRT Question Three Results Tables 117

Appendix F: SRT Question Four Results Tables 124

Casewise Diagnostics 126

Appendix G: Survey and SRT Variables Mapped 131

List of Tables

Table 1: What is teaching? 37

Table 2: How does one improve as a teacher? 38

Table 3: Teaching expertise vs. subject matter expertise 39

Table 4: Variable Information 49

Table 5: SRT Question Text 57

Table 6: Descriptive Statistics for SRT Questions 57

Table 7: Predictor Variables Included in Multiple Regression Analysis 58

Table 8: Descriptive Statistics for Predictor Variables 59

Table 9: Significant Variables By Question 60

Table 10: Model Summary Data for Question One 62

Table 11: Model Summary Data for Question Two 64

Table 12: Model Summary Data for Question Three 66

Table 13: Model Summary Data for Question Four 68

Table 14: Regression Equation Summary Data 69

Table 15: Full SRT Database Average Scores, 2008 74

Table 16: Decile Interpretation of Student Ratings 77

Table 17: Minimums and Maximums of SRT Questions 1-4 78

Table 18: Means and Standard Deviations for Main Effect Variables by Question 101

Table 19: Correlations of Variable of Interest by Question 102

Table 20: Casewise Diagnostics for Question One 104

Table 21: Case Summaries for Question One 105

Table 22: Multiple Regression Coefficients for Q1 107

Table 23: Casewise Diagnostics for Question Two 110

Table 24: Case Summaries for Question Two 111

Table 25: Multiple Regression Coefficients for Q2 114

Table 26: Casewise Diagnostics for Question Three 117

Table 27: Case Summary Data for Question Three 118

Table 28: Multiple Regression Coefficients for Q3 121

Table 29: Casewise Diagnostics for Question Four 124

Table 30: Case Summaries for Question Four 125

Table 31: Multiple Regression Coefficients for Q4 128

Table 32: Variable Map 131

List of Figures

Figure 1. Graph of SRT Database Average Scores, 2008…….75

Chapter One: Introduction

I listen to a lot of undergraduate student complaints on a weekly basis. Many of these complaints are about how boring their classes are. They talk about how the instructor stands at the front of the room and talks at them, rarely actually engaging them with the material. They talk about how much they want to skip class because it does not interest them, or more specifically, the instructor does not seem to understand anything other than the "classic" lecture style. Sometimes these students even point out how much better their high school teachers were, and that they did not realize when they were in high school how good they had it, in terms of engagement with the subject matter and engagement with the instructor. For years now, I have found myself wondering why this happens, and I dream of a world where I never again have to hear the complaint, "That instructor is SO BORING!" In fact, several authors have written about college teaching (Austin, 2002; Boyer, 1990; Carnell, 2007; Erickson, 1984; Haggerty, 2010; Lambert, Tyce, & Featherstone, 2006; Lowman, 1984; Ramsden, 2003; Rust, 2000; Speck, 2003), most of it anecdotal, that confirms this informal impression about the generally uneven level of effectiveness.

It is my belief that graduate programs do not adequately prepare future members of the professoriate for the entirety of their future jobs. In particular, I believe that graduate programs do not include preparation for teaching, which is, for those planning to pursue careers in the professoriate, a critical part of the job. In the following pages, I will expand on this statement of belief by briefly exploring the definition of graduate programs in the United States, and I will summarize what the extant literature says

about what graduate programs are intended to do. Then, I will present a brief review of the literature regarding the job of the professoriate. It should be noted that though the topic may fit here, in order to refrain from repetition, the full literature review on how faculty members learn to teach will be addressed in the second chapter. Finally, I examine the gap between what graduate programs are intended to do and what the job of the professoriate entails, which will lead to a concise research problem statement and associated hypotheses.

Graduate Programs in the United States

The Carnegie Foundation for the Advancement of Teaching began a project in 2001 that was dedicated, in part, to improving doctoral programs across the United States. Termed the Carnegie Initiative on the Doctorate, or CID, one of many proposals it made was that the purpose of doctoral education is "to educate and prepare those to whom we can entrust the vigor, quality, and integrity of the field" (Golde, 2006, p.5). This seems to be a modern and forward-looking definition of the purpose of doctoral education, but what is doctoral education in practice? To answer this, one must look broadly at graduate programs in the United States.

A short, visual survey of doctoral university websites indicates that many universities describe the doctoral degree as preparing students first to perform research. In all cases of the very few websites visited, somewhere later in the description of the doctoral degree appears a reference to educating others, or teaching, or being a faculty member. Even the Council of Graduate Schools (n.d.), in their brochure *Open Doors with a Doctorate*, lists research first and education second. In perusing the public

websites of "Ivy League" doctoral institutions such as Harvard (www.harvard.edu) and the Massachusetts Institute of Technology (www.mit.edu), land-grant comprehensive doctoral institutions such as the University of Minnesota (www.umn.edu) and Carnegie-Mellon University (www.cmu.edu), private single-doctoral-granting institutions such as Gonzaga University (www.gonzaga.edu) and Saint Mary's University of Minnesota (www.smumn.edu), public comprehensive universities such as the University of Georgia (www.uga.edu) and the University of Missouri (www.missouri.edu), and for-profit universities such as the University of Phoenix (www.phoenix.edu) and Argosy University (www.argosy.edu), the doctoral-level degree programs are described as being either research-oriented, or designed to create students who are leaders. It almost seems an afterthought when they add that someone with a doctoral degree can hold a position as a faculty member. While it is true that this list of universities is in no way exhaustive, it does not appear that institutions in the United States that currently have doctoral programs view these programs as more than a vehicle for research preparation or educational leadership.

Literature on Graduate Education

The literature addressing the purpose of graduate education tells a slightly different story. Austin (2002), in particular, asserts that the purpose of graduate education is, in part, socialization for the professoriate. If one approaches graduate education from the standpoint that it is the benchmark that one must meet to get a job within the professoriate, then this assertion makes sense. In 1983, Eble wrote "In almost all graduate programs, the main emphasis is on mastery of the knowledge of a

field and demonstration of research competence in some aspect of that field" (p. 124). Shulman elaborated on this idea during his tenure in the 1980s as the President of the Carnegie Foundation for the Advancement of Teaching, writing that a person with a doctorate "is expected to serve as a steward of her discipline or profession, dedicated to the integrity of its work in the generation, critique, transformation, transmission, and use of its knowledge" (as quoted in Golde, 2006). Refining this idea still further, Gaff, Pruitt-Logan, Sims, and Denecke (2003) wrote the following:

> The hallmark of a doctoral degree has always been and remains the requirement to demonstrate mastery of the field and to apply that knowledge to conduct original research that expands the knowledge base of the discipline. …research has become the dominant or nearly exclusive requirement for the Ph.D. degree, and graduates are well prepared for research-related careers. In addition, all employers expect competencies in other areas, but preparation for such areas is seldom part of doctoral programs. (p. 2)

A graduate degree program is essentially job preparation (Austin, 2002; Bowen & Rudenstine, 1992; Capps, Melvin, Vallano, Weiss, Watkins, Moslemi, et al., 2009; Gaff et al., 2003; Golde & Walker, 2006). As described by the literature, if one accepts that a graduate degree program is intended to prepare the learner for a job in the field, and within that job, primarily to perform research-related tasks as described by the doctoral program websites, what does this mean? More specifically, if one accepts that one of the qualifications for a position in the professoriate is successful completion of the graduate program, what exactly is the job of the professor?

The Job of the Professor

If a graduate degree is the benchmark by which one may attain a job in the professoriate, what exactly does that job entail? The literature does not precisely agree on this. However, Boyer's seminal work, *Scholarship Reconsidered: Priorities of the Professoriate* (1990), stands out in defining what the professoriate is all about. In fact, this is still mentioned today in almost any discussion about the profession of the professoriate (c.f. Austin, 2002; Boshier, 2009; Golde & Walker, 2006; O'Meara & Rice, 2005). Put succinctly, Boyer (1990) suggested that "the work of the professoriate might be thought of as having four separate, yet overlapping, functions" (p. 16). These four functions are the scholarships of discovery, integration, application, and teaching (Boyer, 1990). Boyer's work has been identified as the starting point for not only conversations about, but changes in, the professoriate.

Interestingly, most criticism of Boyer does not state that his four functions fail to define the job of the members of the professoriate; rather, the criticisms tend to attack his definition of each of the four scholarships and the practical application of them. In particular, Boshier (2009) mentions that there is a great deal of confusion on what the word 'scholarship' actually means, because it gets used for so many different things, and the difficulty in operationalizing Boyer's definitions results in marginalization of the entire work. Regardless, many have published on the topic of Boyer's definition of the work of faculty: scholarship.

Beyond Boyer's work and the numerous references and responses to it from other authors, there is a substantial amount of literature that also describes the job of the

professoriate, from the standpoint of the first- year faculty member. More specifically, these books and articles offer advice on how to succeed during one's first year and further as a faculty member (c.f. Eble, 1972; Eble, 1983; Lowman, 1984; McKeachie & Svinicki, 2006; Parker, 2002; Pickering, 2006; Ramsden, 2003; Ross, 1986; Rust, 2000). The point here is not that training does not exist for new faculty members. The point is, instead, that the majority of new faculty members receive training on the job instead of prior to the job. This begs the question about the minority of faculty that receive training prior to the job.

Preparing Future Faculty Programs

Though the majority of new faculty members receive training after they have been hired, some have been fortunate enough to participate in Preparing Future Faculty programs (PFF). The full literature review on this program appears in the next chapter. Here, however, it is important to understand of what these programs consist. I will take as my example the well-regarded PFF program at the University of Minnesota. According to the literature published by the Council of Graduate Schools and the Association of American Colleges and Universities, the PFF programs nation-wide were created because of the assertion that "doctoral students planning to join the faculty should learn about each of these elements [teaching, research, and service] of the academic profession prior to earning the degree" (Gaff et al., 2003, p. v).

The PFF program at the University of Minnesota began in 1993 with the funding of the PFF program nationwide by the Pew Charitable Trusts (Gaff et al., 2003; Preparing Future Faculty, n.d.b). Now, 17 years later, the program is sponsored by the

UMN Graduate School and the Office of Human Resources, and it consists of two, three-credit-bearing classes that graduate students may take. When finished, these students earn a certificate. The classes cover topics related to the scholarship of the professoriate, including learning theory, teaching skills, job search strategies, instructional design, collegiality, and institutional culture and type (Center for Teaching and Learning, n.d.).

Having been a participant in this program at the University of Minnesota, it is my firm belief that such a program is extraordinarily valuable for any graduate student who plans to pursue a career in the professoriate. It is within the auspices of this program that I have come to a more comprehensive understanding of what the career entails; in my case, at least, the program has met its goal. However, it is important to note that my participation was purely voluntary. It was not a required part of my degree program. In fact, in informal discussion with the other students in the classes I have taken, I discovered that some of them were required to take classes in the program not as part of their degree programs, but in order to retain jobs as teaching assistants (TAs) or post-doctoral graduate assistants (GAs) (GRAD 8101 and GRAD 8102, personal communication, 2009-2010). In their cases, the coursework was somewhat constrained by the need to earn money, yet was still at least partially voluntary: they could have chosen to fund their study through other means than TA or GA appointments and avoided the classes altogether.

It is possible for students who plan to pursue careers in the professoriate to find ways to acquire teaching skills. The above description of the PFF program is one such

way. Yet, it is my contention that teaching skills should be directly addressed in graduate programs for these students, and they are not. Providing access to such a program is not the same as making preparation for teaching a requirement in order to earn the degree.

Researchable Problem Statement

We do not know whether there is a relationship between how faculty members learn to teach and how effective they are at teaching undergraduates, as judged by student evaluations of their courses. As mentioned above, this research problem statement adds in a new element, that of student evaluations. In the next chapter, I will discuss course evaluations, known in the literature as student ratings of teaching, or SRTs. This question is designed to fill a gap in the research regarding preparation for teaching and the results of that preparation. The purpose of filling this gap, if there is a positive relationship, is to provide quantitative evidence to support a change in graduate education to include instruction in teaching for those graduate students who plan to pursue careers in the professoriate.

Need

This study is needed because currently new faculty members generally do not have any teacher preparation, and struggle to learn "on the fly" to teach, either by having a mentor, by attending learning sessions at the college or university's center for teaching and learning, or by remembering how their professors taught. Evidence of the existence of the struggle to learn to teach is well-documented in Preparing Future Faculty programs across the United States, as well as in literature exhorting professors

to improve their teaching skills and giving them advice on how to do that. However, there is little quantitative data on how new faculty members learn to teach, and this study will explore what those ways are, and whether there is any relationship between how they learned to teach and their effectiveness as teachers.

Hypotheses

The null hypothesis is that there is no relationship between how faculty members learn to teach and their effectiveness as determined by course evaluations. The alternative to the null hypothesis is that there is a positive relationship between how faculty members learn to teach and their effectiveness as determined by course evaluations.

Conclusion

It may seem like a big jump from the opening paragraph of this chapter to go from "Why is it that instructors of undergraduate students do not know how to teach?" to "Is there a relationship between how faculty members learn to teach and how effective they are at teaching?," but in all actuality it is a very small jump. There are many reasons why instructors of undergraduate students may not know how to teach. There is also a great deal of literature, which will be reviewed in the next chapter, about helping those instructors to do a better job.

What seems to be missing from the mix is understanding whether learning to teach before getting the teaching job actually helps faculty members be more effective at teaching. Logic would claim that a positive relationship exists between these two

factors. However, there is no quantitative data proving this claim, which results in a gap in the research. This paper will serve, at least in part, to fill that gap.

Chapter Two: Literature Review

The purpose of this chapter is to examine the literature on two topics: how faculty members learn to teach, and teaching effectiveness. In order to give perspective on how the United States stands out regarding how faculty members learn to teach, this topic will also be explored in other countries that have systems of graduate education.

The first part of this chapter will review the literature that discusses how faculty members learn to teach. The second section of this chapter will explore the topic in other countries. The third section of this chapter will address teaching effectiveness. Finally, the fourth part of this chapter will discuss the relationship between the two elements: how faculty members learn to teach, and teaching effectiveness.

How Faculty Learn to Teach

How do faculty members (including graduate students and adjunct faculty) learn to teach, and what effect does this have on the learners? Initially, I thought that the majority of faculty members, broadly defined, learn to teach through mimicry and trial-and-error methods rather than having formal training on how to teach as part of their degree program, and that this adversely affects student learning. After doing research on the topic, I found that there is little published research on how faculty members learn to teach. Even less has been published about when faculty members learn to teach – before or after acquiring the degree that has allowed them to become faculty members.

Also appropriate for this work is questioning what anyone needs to know in order to engage in the professional role called “teaching”. This will be discussed below in more detail.

What Are Teaching and Learning?

In the absence of a wealth of research and resultant literature on how faculty members learn to teach, I turn to the question of what teaching and learning are. Ramsden (2003) said it the most simply:

> The aim of teaching is simple: it is to make student learning possible. Teaching always involves attempts to alter students' understanding, so that they can begin to conceptualize phenomena and ideas in the way … that we as academics want them to understand them. (p. 7)

Boyer (1990) agrees and extends the definition: "teaching, at its best, means not only transmitting knowledge, but transforming and extending it as well" (p. 24). Others present definitions that fall in between these two, resembling both (cf. Brookfield, 1990; Eble, 1983; Speck, 2003). I prefer Boyer's definition because it is very concise. It does force one to think further about what one needs to know in order to do this, and one of those things is assumed to be understanding what learning means.

There is a wealth of literature that describes the nature of learning, and that attempts to define learning. Ramsden defined learning very simply as a qualitative change in a person's view of reality (Ramsden, 2003). Other interpretations include the acquisition of information, knowledge, understanding, or wisdom (Ackoff, 1996); a process that results in changes in knowledge, beliefs, and behaviors (Watkins & Marsick, 1993); and that it is the result of teaching (Trigwell & Shale, 2004). There are others, and these few give an idea of the breadth of research available on what learning actually is. I prefer Ramsden's definition due to its elegance.

Assumptions that faculty can teach. "It is assumed that all faculty can teach, and hence that one doesn't need to spend a lot of time on it" (Boyer, 1990, p. 32). This is assumed because faculty members are subject matter experts. This is a key assumption in academia about what one needs to know in order to be able to teach. The doctoral or master's degree conferred upon the people holding positions as members of the faculty is the guarantor of subject matter expertise: "the scholarship of a particular discipline, and increasingly a specialization within the discipline…has been seen as the essence of graduate school education" (Speck, 2003, p. 42). Further, it is assumed that teaching is simply conveying knowledge (Speck, 2003). The break in this chain of logic is that teaching is not simple: it is much more than conveying knowledge. Among other things, it is something that one must be taught how to do. There is a wealth of literature on teacher preparation for K-12 that supports the idea that one must be taught how to teach. One simply has to look at the undergraduate curriculum for any college or university that offers a degree in education, or at each state's qualifications for an individual to earn a teaching license to find evidence for this. This evidence shows that, in addition to subject-matter knowledge, in general, there are things determined by the college, university, or state legislators about what one needs to know in order to be qualified to teach. The idea that one can learn to teach through osmosis is bankrupt: "Teachers are apparently supposed to know how to teach because they have been watching teachers do it since first grade—kind of like learning how to play tennis by sitting in the grandstand" (Huber, 1992, p. 124). As with most skills, one must have the opportunity to learn the skill and then the opportunity to practice it to become

proficient. It is the unusual and rare person who can learn only by watching someone else, and no learner truly learns without the opportunity to practice. Memory, which is what one has in the absence of practice, is not the same as learning.

Brookfield (1990) addresses another insidious myth regarding faculty and teaching:

> There is a dangerous myth abroad that the best teachers are those who were the most successful as learners. According to this myth, someone with a degree from a prestigious university and an A+ average will be better equipped to teach a subject than someone with a B average and a degree from an underfinanced state college. This myth correlates intellectual achievement with pedagogic expertise, and it underlies the efforts to make teaching an all-graduate profession. It also underscores the system of incremental payments in some institutions whereby attaining a master's degree, then a doctorate, results in the teacher moving one step up the pay scale. (pp. 8-9)

This myth seems very similar to the "if they know it, they can teach it" theory, perhaps as a variation on the theme. Brookfield's comment on pay and the relationship of degree earned to pay received in the professoriate is interesting, and it is more pertinent to the K-12 teacher than the post-secondary-level teacher, primarily because the post-secondary-level teacher generally must have the terminal degree in the field just to get a job in academe. One would think, though, that if the best teachers were the students who were the best in school, the professoriate would remain an elitist institution, drawing only the top graduates from the top schools.

What other types of teachers know. The programs that K-12 school teachers must go through to become certified is evidence that a teacher must be taught how to teach. Different states require different things for certification, and so I turn to the NCATE (National Council for the Accreditation of Teacher Education) website for their standards:

Teacher candidates reflect a thorough understanding of pedagogical content knowledge delineated in professional, state, and institutional standards. They have in-depth understanding of the subject matter that they plan to teach, allowing them to provide multiple explanations and instructional strategies so that all students learn. They present the content to students in challenging, clear, and compelling ways and integrate technology appropriately. They develop meaningful learning experiences to facilitate learning for all students. They reflect on their practice and make necessary adjustments to enhance student learning. They know how students learn and how to make ideas accessible to them. They consider school, family, and community contexts in connecting concepts to students' prior experience and applying the ideas to real-world problems. (2006)

If these are the things that teachers need to know how to do to teach a student who is a senior in high school, why should a faculty member at a college or university not know the same ways to teach that student when he or she arrives on campus after high school graduation, when he or she is a scant three months older? Unfortunately, that standard is not being universally applied, perhaps because there is no oversight for colleges and universities as there is for K-12 schools.

The brief literature review above indicates that teacher training is not generally a part of graduate degree programs. If we accept that faculty members are not taught how to teach during their graduate programs, the next question that must be answered is when do they learn to teach? There is a great deal of literature regarding how new faculty need help with all aspects of teaching: syllabus construction, organizing a classroom, instructional presentation, and so on (cf. Brookfield, 1990; Eble, 1983; Lowman, 1984). The fact that this wealth of literature exists can be taken as evidence that faculty members either learn by doing, which may also include getting feedback from a mentor while teaching, or learn by reading the literature, reflecting, and then doing. In other words, either they are reading the self-help books and putting what they read into practice, or they are attending classes on their own.

Developing better teachers. There is ample research, as well as theory, on what elements make up the professoriate. Boyer (1990) wrote, "the work of the professoriate might be thought of as having four separate, yet overlapping functions. These are: the scholarship of discovery; the scholarship of integration; the scholarship of application; and the scholarship of teaching" (p. 16). Eble and McKeachie (1985) wrote about faculty development and identified similar elements. They also did a thorough literature review of faculty development, which closely matches the areas identified by Boyer five years later.

Not only are faculty able to use campus resources to develop themselves (Boyer, 1990), but there are numerous books published about how faculty can get better at what they do. Considered a seminal work, *Learning to Teach in Higher Education*

(Ramsden, 2003) covers many elements of how faculty might do better jobs as faculty members:

> To achieve change in the quality of teaching and learning, we ought rather to look carefully at the environment in which a lecturer works and the system of ideas which that environment represents. This means an emphasis on team curricula, courses and departments, as well as on individual academics. (p. 9)

Ramsden goes on to comment on what makes up the knowledge base for all teaching in higher education: "[a] deep understanding of learning and teaching and their relationship to each other is an essential base for effective action as a university teacher" (p. 11). Other authors (Brookfield, 1990; Carnell, 2007; Dall'Alba, 2005; Erickson, 1984; Lowman, 1984; Pill, 2005) have laid out programs or research to help faculty members become better teachers. McKeachie and Svinicki (2006) edited McKeachie's ubiquitous book, *Teaching Tips: Strategies, Research, and Theory for college and University Teachers*, and this work is widely used both in Preparing Future Faculty programs and as reference material for teaching assistants across the United States. All of these authors propose ways that faculty members can improve.

Research vs. Teaching

There is a substantial amount of literature on the relationship between research and teaching. Brew (2003), Deem and Lucas (2006), and Hattie and Marsh (1996) wrote on this topic, and in doing so, surveyed a substantial amount of the extant literature. Brew's literature review can be summarized this way: "research and teaching are both viewed as activities where individuals and groups negotiate meanings, building

knowledge within a social context" (2003, p. 12). Deem and Lucas cite Hattie and Marsh, who say that there are three negative-, two positive-, and three zero-relationship theories about the relationship between teaching and research (cf. Hattie & Marsh, 1996, and Deem & Lucas, 2006). Deem and Lucas' literature review may be summed up in that there are four ways that teaching and research can be linked: the transmission model, the cultural model, the learning model, and the critical inquiry/symbiosis model (Deem & Lucas, 2006, p. 7).

There is also a substantial amount of literature on how research and teaching affect each other. The way these affect each other seem to be obvious and simple to accomplish, but the research does not support that assumption (Lindsay, Breen, & Jenkins, 2002). Some say that the skills needed to be a competent teacher are not related at all to those needed to be a competent researcher (McIntyre, 1997). Ramsden (2003), on the other hand, says that that students enjoy being taught by competent researchers, which would indicate that the knowledge generated by research enhances teaching. Still others, most notably McLean and Barker (2004), argue that some research activity on the part of university teachers is a strong—though not necessary—'pre-condition' of desirable educational goals. There is no universally agreed-upon definition or model that explains the relationship between teaching and research.

An Exploration of the Professoriate in Europe and the UK

The basic premise of this dissertation is to discover whether there is a relationship between faculty teaching preparation and student ratings of teaching. To that end, one element of this exploration will need to include the exploration of

certification, or qualification to perform a job. This section will specifically explore the professoriate in Europe. It has been most revealing to research this topic, as it would appear that the qualifications to be a professor are not as much under scrutinyin Europe as they seem to be in the United States. Regardless, in this section I explore how professors are trained in Europe. I will explore the Bologna Declaration of 1999, and will discuss research regarding the implications of this declaration and more recent European agreements that result from this declaration. Then, I will explore the concept of quality teaching in Europe, and will conclude with a discussion of how quality is related to the Bologna Declaration and thus to teaching in the European professoriate.

The Professoriate in Europe

I believe that documented preparation to perform a job of teacher, broadly defined, should be part of a person's statement of qualifications for the job. At one time, I believed this to be true across the board, without regard to content area or locale. I have discovered, upon researching the concept in Europe, that this belief is extremely American-centric. It does not work as well in the context of the European professoriate.

There is a very different view in Europe of graduate school and how one earns a graduate degree. This is important, because in America, for the vast majority of positions in the professoriate, one must have a terminal degree, and that terminal degree is usually a doctoral degree. This section will discuss the Bologna Declaration of 1999 and introduce and summarize the recent work of the European University Association. For a history of the movements in Europe that led to the Bologna Declaration, see Barblan (2002).

The Bologna Declaration of 1999. The Bologna Declaration seems to be the basis for many of the reforms of higher education in Europe. It was the product of collaboration between the European Ministers of Education, and was signed by 31 of same. It was built on several previous documents and meetings, and there is ample evidence on the world-wide web to support this. The Declaration itself mentions the Bologna Magna Charta Universitatum of 1988 and the Sorbonne Declaration of 1998, both of which were products of the meetings of organizations that were precursors of the European University Association.

The importance of the Bologna Declaration lies in the recognition of two factors. First, the Bologna Declaration firmly states that all the signatories acknowledge and recognize "the importance of education and educational co-operation in the development and strengthening of stable, peaceful and democratic societies" (p. 1). Second, it lays out specific agreements that are foundational to all future work of the participant countries with regard to higher education. These include adopting a system of easily readable and comparable degrees, adopting two main cycles (undergraduate and graduate), establishing a credit system that would promote student mobility, promoting that mobility, promoting cooperation in quality assurance, and promoting needed aspects of higher education, most notably in curriculum development, institutional cooperation, mobility, and training/research (The Bologna Declaration, 1999).

This initial agreement cleared the way for several future meetings and agreements pursuant to the ends laid out in the Bologna Declaration. Some of these

include the Lisbon Summit of 2000, also referred to as the Lisbon Strategy (Kehm, 2007), the 2003 Berlin Communiqué which established doctoral education as the third cycle (Rapp, 2008), the 2005 Salzberg Principles, the 2006 Nice Conference, and the 2007 London Communiqué (Chambaz, 2008). By mentioning these, it is not the intention to explore each one. There is ample information available from the European University Association website, and also through "googling" any of these meetings and agreements. Instead, it is important to acknowledge that each of these built on the previous one, and all seem to take their beginnings from the content of the Bologna Declaration. Further, it is important to note that the European University Association seems to be built on the premises of cooperation laid out in the Bologna Declaration of 1999.

The European University Association. The European University Association (EUA) was formed in 2001 by the merging of the Association of European Universities and the Confederation of European Union Rectors' Conferences (EUA, n.d.). The EUA supports higher education institutions in 46 countries, and being a part of the European Union neither promotes nor denies membership in this association. In other words, this is an educational organization that, while it deals with national and supranational politics, is not itself a political entity.

The EUA holds a yearly conference, and one of the more recent conferences, in 2008, launched a Council for Doctoral Education (CDE). The CDE takes as its mission "to contribute to the development, advancement and improvement of doctoral education and research training in Europe" (EUA, n.d.). It is this mission that becomes most

critical to the research uncovered in support of this section. Kehm (2007) wrote about the changes that the EUA has been promoting in the entities of the Bologna Declaration and Lisbon Strategy, and pointed out that changes in doctoral education and research training have moved from being institutionally-led to nationally- and even supranationally-led. It seems that the CDE's very existence supports her conclusions.

It is further interesting to note that the EUA's yearly conferences have produced the 2003 Berlin Communiqué, the 2005 Salzberg Principles, the 2006 Nice Conference, and the 2007 London Communiqué. As previously mentioned, the Berlin Communiqué established agreement among the participating organizations that the Bologna Declaration needed to be expanded: there was a third cycle, commonly referred to as the doctoral education cycle. The Salzberg Principles indicated that the members agreed that the "core component of doctoral training is the advancement of knowledge through original research" (Chambaz, 2008, p. 5). The 2005 Nice Conference established a group to prepare recommendations on doctoral programs for the 2007 conference in London, and the London Communiqué was simply the results of those recommendations, written in a report commonly titled TRENDS V. This was a report on the implementation of the Bologna Declaration.

The EUA continues its work today in driving change in European education. The CDE, as its most recent working group, seems to be concentrating on the mobility of doctoral education throughout Europe, as well as gaining widespread agreement with the Bologna Declaration (EUA website). It has made some significant changes in the ways European universities view graduate education.

European Universities and Graduate Degree Programs

At this point it is important to look at how the Bologna Declaration and the EUA have affected European Universities. First, however, it is necessary to understand how European universities have traditionally offered graduate degree programs, and then to explore the literature regarding a select few European countries.

Ways European universities offer graduate degree programs. Huisman and Naidoo (2006) provide a very short history of the doctoral degree, commonly referred to as a Ph.D., Doctor of Philosophy, or Philosophiae Doctor in the original Latin. They claim that the "traditional" Ph.D. is understood universally simply because it was a product of the global diffusion of the European medieval university model, and was adjusted based on the culture and practice of the higher education systems to which it migrated. Huisman and Naidoo contend that this universal understanding is that 1) the Ph.D. is the highest degree awarded in higher education, and 2) the doctoral candidate "should make an original (and significant) contribution to academic knowledge" (Huisman & Naidoo, 2006, p.4).

According to a survey administered by the EUA in 2006, there are four main ways that European universities offer graduate degree programs. Chambaz (2008) lists these as individual-based, structured programs, doctoral/graduate/research schools, and combinations of these. An individual-based program is equivalent to an apprenticeship program, with one professor imparting knowledge to one student in whatever way the professor feels is best. As Bitusikova (2009) wrote, "doctoral education followed a model of individual study (known also as the "master-slave" or the apprenticeship

model) that was based on a working alliance between the doctoral candidate and the supervisor and did not include a structured coursework phase" (p. 2). She goes on to write that a structured program will include specific coursework and phases, and a doctoral/graduate/research school may be organized around a university or a conglomerate of universities, and has "clear administration and leadership and specific funding support" (Bitusikova, 2009, p. 2).

It is interesting to note that doctoral education and research training are treated as parts of the same entity by the EUA. Publications suggest that doctoral research in Europe is seen as "the first formative stage of a research career" (Rapp, 2008, p. 3), and further, that "doctoral education & research careers have become a crucial element of national and European competitiveness strategies" (Rapp, 2008, p. 2). The CDE is seen as "the platform for European universities to discuss, develop and implement good practices in doctoral education leading to high level research and training and enhancing attractiveness and competitiveness of universities" (Rapp, 2008, p. 12). It is further interesting to note that, while training is listed as an element throughout these declarations, agreements, mission statements, and publications, it is very unclear what is meant by training. This will be explored further below.

Quality Assurance and the European Professoriate

In completing research for this section, it was discovered that many of the authors of peer-reviewed journal articles seem to equate quality assurance in the professoriate with an individual professor's teaching. There are, of course, some institution-wide levels of teaching quality assurance that have been studied (c.f. Koler &

Huber, 2006; Romainville, 1999), but for the most part these seem to include some element of assessment of teaching.

This section will first explore teaching as an element of the professoriate. The Czech Republic and Germany are explored. It should be explicitly noted that there are a very limited number of articles on other European countries' systems of higher education as they relate to either teaching as part of the professoriate, or to the Bologna Declaration. This, along with space and time limitations, is why only the Czech Republic and Germany will be explored. Next, the concept of quality assurance is briefly examined in the greater context of the European professoriate. Finally, some conclusions are drawn regarding how quality assurance is related to the European professoriate and the Bologna Declaration.

The Case of the Czech Republic

Vasutova (1998) recognizes both that teaching is a critical element of the professoriate, and that it is highly under-valued in the Czech Republic. Vasutova (1998) goes on to list several elements of teaching, including curriculum development, actual instruction, student achievement, and authorship of textbooks and teaching materials, and to lament the fact that, at least in the Czech Republic, the quality of teaching is not monitored sufficiently.

It is interesting to note that in the Czech Republic, teaching competence is a highly sensitive issue, with many in the profession affirming that "pedagogical competence in higher education is not worth considering, and that it is the responsibility of the students to deal with their study requirements" (Vasutova, 1998, p. 478-479).

Further, the Czech professoriate seems to believe that instruction can be reduced to the delivery of information and the reproduction of same by the students (Vasutova, 1998). It should be noted that this is a key question: whether learning is the result of teaching quality or of the characteristics or effort of the student. The Czech professoriate seems to believe it is the latter.

The Czech professoriate consists of four levels, determined by the higher education laws of 1998. These levels include instructor, assistant professor, associate professor, and professor. Movement from assistant to associate professor is determined by an assessment of habilitation, which is based on his or her teaching experience, and the quality of that teaching experience. There are, of course, other elements that are assessed (previous training, a public habilitation lecture, field work, and fellowships), but it is most notable that Czech law determines that one of the qualifications for the title is teaching experience.

German Universities and Doctoral Education

Much has been written on the German system of education (c.f. Winch, 2006), and so it is not the intention to cover this information again here. Huffner and Landfried (2003) note that "the conviction that the quality of teaching in all German higher education institutions is of the same high standard everywhere is no longer valid" (p. 141). This is important because it assists the understanding that Germany is in a different place than the Czech Republic regarding higher education. Much time and effort is being spent on ensuring that German higher education is competitive with the rest of Europe. In contrast, Czech higher education is concentrating on implementing

the Bologna Declaration. Huffner and Landfired (2003) in particular state that the quality of teaching in German higher education is going to drive changes in the near future.

Germany views itself as a leader in shaping higher education markets (Hahn, 2003). However, while much effort is being spent on making German higher education competitive with the rest of Europe, the framework for these changes is the implementation of the Bologna Declaration. This is seen in the contrast between the traditional one-cycle German system of education and the three-cycle system of the Bologna Declaration. Guth (2006) supports this assertion: "Increasing the efficiency and attractiveness of German higher education as a whole could therefore make a significant difference … by reducing the time to degree and attracting foreign researchers to Germany" (p. 329). Germany, in order to become competitive with Europe, must align itself with the three-cycle system and increase mobility of students into and out of its higher education system (Guth, 2006).

The reader should note that there is very limited information on the status or view of teaching in German higher education. This is an area that perhaps needs more research. As mentioned above, the Bologna Declaration specifically calls out training, which seems to mean training as a teacher. However, in researching German higher education in conjunction with the Bologna Declaration, it became very clear that teaching was not the primary, nor even the secondary nor tertiary, concern for Germany. Instead, competitiveness with the rest of Europe, probably caused by differing

viewpoints on the number of cycles of education and what doctoral education consists of, seems to be German higher education's primary concern.

Quality Assurance in Higher Education

Quality assurance in higher education covers a lot of territory. At an institutional level in Europe, classic quality assurance has four parts: self-evaluation, the construction of a team of external experts, the evaluation activities of that team, and the report produced by that team (Romainville, 1999). However, it is much more controversial to discuss the quality of teaching, and, of course, there is a great deal more literature on this topic. It is perhaps more controversial because quality is much less measurable in teaching, and yet, it is critical to have high-quality teaching in higher education.

Defining teaching quality is also a topic subject to much controversy. Many have written on the topic (c.f. Eble, 1983; Brookfield, 1990; Dall'Alba, 2005). Leitner (1998) wrote of a specifically German model, arguing that teaching quality involves certain quasi-universal aspects, adjusted to meet the needs of the university and culture. These include teacher preparation, structure of the course, pedagogical professionalism, clarity of presentation, media use, and student-instructor interaction. He goes on to conclude that every teacher in higher education should be given the opportunity to improve his or her pedagogical skills, if for no other reason than that the teacher may some day be evaluated on them. Romainville (1999) refines quality assurance in higher education further, concluding that the list of conditions may be utopian in nature in current European practice, but are still worthy. Specifically, Romainville (1999)

recommends that high-quality teaching be student-centered in many ways, including inviting students to become more involved in and responsible for their own learning. This is certainly a departure from the Czeck perspective that learning is the student's responsibility.

In order to bring this information back full-circle, it is useful to explore Kohler's (2003) article on quality assurance and the relationship with the Bologna Declaration. He points out, correctly, that the Bologna Declaration and all subsequent communiqués (at that time) do not define quality or quality assurance. Instead, quality is left up to the member institutions to define. Kohler (2003) gives six considerations that build on each other, but only the first three need to be considered here. The second three are particular to institutions, and this dissertation is primarily concerned with the professoriate, and with teaching as an element therein. The first consideration contains multiple parts, and again we need be concerned with only one of them: academic quality, which also covers excellence in teaching. The second consideration is that quality should be based on the outcome of the learning experience. Again, this can be viewed as directly related to teaching. The third consideration regards standardization as it is related to the Bologna Declaration. It specifically says that quality is, and should be, relative, and standardization is only possible to ensure mobility. More specifically, the Bologna Declaration gives a framework for making European higher education transferable across country lines, and standardization comes into play in order to be able to compare higher education across those country lines. Another way to look at this is in from the viewpoint of the German system, previously discussed. In order for the

German system to become competitive with the rest of Europe (as stated in the research, above), the German system needs to adopt some elements of standardization from the Bologna Declaration, and therein demonstrate the quality of the German system as compared to other European systems.

The Spellings Commission

On September 19, 2005, U.S. Secretary of Education Margaret Spellings announced that she had commissioned a group of distinguished educators, CEOs, and policy makers to look at the future of higher education in America (US Department of Education, 2005). The intent of the commission was to explore the questions of what skills would be needed in the 21st century, how to keep the US in a leading role in higher education, and making higher education accessible to all. Note the focus on outcomes in the American commission. The 19-member group spent just under a year engaging in dialogue with various stakeholders regarding these questions. In June, 2006, the Spellings Commission published their final report.

The Spellings Commission reported on only a few findings regarding teaching. First, it found that the results of research on teaching were not being put into practice, regardless of the level of the student. This seems to be a harsh condemnation of the state of teaching in higher education. The final report went on to state that technological advances in teaching should be supported by policymakers at the state and federal levels, and that colleges and universities should work with K-12 teachers to ensure that they know how to teach (US Department of Education, 2006). There were other statements about teaching, but these did not apply to higher education, or applied

only to K-12 teachers who were enrolled in higher education, not to those who teach in higher education.

Reactions to the State of Teaching in Higher Education

Some graduate schools require their students to demonstrate teaching ability prior to earning the doctoral degree, while other schools do not allow their Ph.D. candidates to teach undergraduate classes (Douglas, 1997). I attended one such institution. As an undergraduate, the only graduate student I had as an instructor was one who was in the process of defending his dissertation. All of my other classes were taught by full-time faculty members. Even the Nobel-prize-winning physicists and economists were required to teach as part of their contract with the university; there was no such thing as a research-based appointment. However, I recognize that the institution I attended as an undergraduate was unusual, and perhaps even unique, in that respect. I have found it odd that there has not been a recognition and acceptance by academe of the need for "an inquiry into the relevance of the skills learned in graduate school to the tasks professors are actually required to carry out, especially in those institutions where research is not emphasized, might shed light on the utility of the doctorate as a teaching credential" (Douglas, 1997, p. 149). This leads us to the Preparing Future Faculty program.

Preparing Future Faculty programs. Launched in 1993 and funded through 2003, the Preparing Future Faculty (PFF) program was a cooperative effort between the Council of Graduate Schools and the Association of American Colleges and Universities. Its intent was to address the full scope of professorial activity, defined by

PFF as including teaching, research, and service (Preparing Future Faculty, n.d.a). The literature from the organization explicitly recognizes the space the program is intended to fill: "The degree does not prepare these highly skilled research professionals to be faculty members. Academic employers increasingly also expect new faculty to be excellent teachers" (Preparing Future Faculty, n.d.a).

This program enjoys considerable success at the 88 involved institutions. DeNeef (2002) writes that PFF alumni have found participation in the program to have helped them tremendously in their academic careers. In fact, DeNeef's article states that the PFF program goes a long way towards helping graduate students overcome the prevailing faculty viewpoint that learning about teaching is not something that is important in graduate studies: "Since many of them [faculty members] do not think of themselves primarily as teachers, they are not always appreciative of the instructional needs of their graduate trainees" (DeNeef, 2002, p. 7).

To give a more specific example, at the University of Minnesota, the Preparing Future Faculty program is somewhat more intense than at many other universities. There are several for-credit, graduate-level courses offered. These courses, and indeed the program as a whole, concentrate on developing teaching skills, classroom and job search materials, and mentoring relationships with faculty at various types of institutions. The existence of this program indicates that the University of Minnesota believes in the value of teacher preparation. The funding of the program by the Graduate School indicates that the concept of pedagogical training for graduate students is valued as part of their graduate career. The continued existence of the program 17

years after it began and 7 years after it was no longer funded by grants indicates that the University and the program participants find the program valuable.

The existence of programs like the PFF program is a bright light in an otherwise grey and gloomy sky of teaching in higher education. Brint (2008) suggests that teachers in higher education should professionalize their practice before external regulators do so. Brint uses as a starting point the Spellings Commission report and moves from there to its successes as a result of the current emphasis in America of accountability in outcomes. The warning is clear: quality teaching is not currently part of academe, and it should be, because there is a "growing opposition of much of the American public to continuing "business as usual" in higher education" (Brint, 2008, p. 2). Katz (2006) makes it clear that members of the professoriate must teach: "each of us, in whatever segment of the professoriate, has a duty to teach" (Katz, 2006, p. 5). He cites as an example the horror the investigating committee felt at discovering that graduate students generally do not receive teacher training as part of their degree programs. Austin and McDaniels take this horror one step further and cite the following: "Graduate students also report that some advisers urge them to avoid spending time on seminars or workshops pertaining to teaching, since participation in these activities may diminish the available time to spend in the laboratory or on other research activities" (2002, p. 55).

Overall, scholars seem very pessimistic about the current and future states of teaching in higher education in the United States. Multiple authors have cited the necessity of changing this downward trend. Some have even attempted to use a fear

tactic by saying members of the professoriate should police themselves before outside sources attempt it and create unacceptable limits on other aspects of academe.

Teaching Effectiveness

Teaching effectiveness is a much-discussed topic in higher education. There is no agreement on the definition of teaching effectiveness (Stark-Wroblewski, Ahrling, & Brill, 2007), despite reams of research attempting to define teaching effectiveness, determining what it should measure, and discussing how the results should be used. In measuring teaching effectiveness, student ratings of teaching (SRTs) are typically used (Aleamoni, 1999; Shao, Anderson, & Newsome, 2007; Sweeney, Morrison, Jarratt, & Heffernan, 2009). These are also referred to as student evaluations of teaching, much to the consternation of professionals in the field of evaluation (cf. Arreola, 2007). For the purposes of this paper, I will refer to student ratings of teaching, or SRTs.

There is a large volume of published literature that speaks both for and against SRTs as this typical measure of teaching effectiveness. Marsh (1987) wrote a monograph that is frequently cited, and generally treated as foundational literature regarding student ratings of teaching. Overall, he wrote that SRTs are used for a number of reasons by students, faculty, and administrators; are multi-dimensional, reliable, stable, and relatively valid; and there is more work to be done on them. For reasons including availability, ease of use, candor, and limitations of other approaches, student evaluations are necessary to evaluating teaching effectiveness (Jackson et al., 1999). SRTs are said to be valid because they meet convergent validity requirements in that they are congruent with "supervisor, colleague, alumni and observer ratings" (Beran &

Violato, 2005, p. 594). Moore and Kuol (2005) suggested that SRTs can be used in a teaching improvement effort.

On the other hand, SRTs are influenced by several sources of bias, including student interest, motivation, and expectations, as well as teacher behaviors, likeability, expressiveness, and charisma (Olivares, 2003). There is evidence for students being able to distinguish between engaging teachers and effective teachers (Johnson, 2000), and evidence saying that students cannot make this distinction (Moore & Kuol, 2005). There is some question about whether SRTs are used for improving teaching, or whether they are simply a vehicle to meet the bureaucratic needs of the institution (Johnson, 2000; Moore & Kuol, 2005).

The more recent literature presents several concise conclusions about SRTs. First, though there is no agreement yet on the definition of teaching effectiveness, SRTs continue as the standard method for the evaluation of teaching effectiveness in higher education (Moore & Kuol, 2007; Shao, et al., 2007). Second, it may benefit teachers in higher education to add additional approaches for assessing quality, such as pre- and post-tests, teaching portfolios, and peer assessments of teaching (Arthur, Tubré, Paul, & Edens, 2003; Stark-Wroblewski, et al., 2007). Third, SRT scores are higher in institutions where teaching and research are both valued than they are in institutions where one is valued more highly over the other (Terpstra & Honoree, 2008).

Conceptual Framework

Research is made better if a conceptual framework supports it. This section will outline such a framework, exploring different authors' concepts in answer to the

questions that frame the concepts. In order to teach, one ought to know one's subject matter, and one ought to know how to teach. Some initial questions that go into this framework are these: what does it mean to know how to teach? What does it mean to know one's subject? Is knowing how to teach separate from subject matter?

From the competency-based, K-12 teaching model, there is a framework for learning how to teach. How one learns that model within higher education is still up for grabs. It may be that people who want to teach are bright enough to learn the model on the job. Still, the question of how higher education instructors learn to teach does not have any quantitative support yet, and one must keep in mind that the effects of what teachers do sometimes do not really impact people until much later.

What is teaching? Several authors define teaching in different ways. Most give a definition that combines knowledge with student learning goals. Also, the authors quoted tend to be concerned with effective teaching, instead of general teaching.

Table 1

What is teaching?

Author	Evidence
Brookfield, S. (1990)	It is a myth that the most effective teachers are the ones who were most successful as students.
Knight, P. T. (1998)	Teaching is related to a broad range of learning goals.
Ackerlind, G. S. (2003)	It is necessary to understand what teaching is in order to improve.
Brew, A. (2003)	"… research and teaching are both viewed as activities where individuals and groups negotiate meanings, building knowledge within a social context" (p. 12).
Trigwell, K. & Shale, S. (2004)	Effective teaching results in student learning.
Hutchings & Schulman, as quoted in Austin, A. E., & McDaniels, M. (2006)	Effective teaching results in faculty-student engagement and fosters learning in the classroom.

How does one improve as a teacher? This question seems to be answered from the standpoint of how one learns to teach and the standpoint of once one knows how to teach, this is how one gets better at it. Interestingly, most of the authors indicate that skills practice is a part of learning about teaching practice.

Table 2

How does one improve as a teacher?

Author	Evidence
Parker, J. (2002)	Teaching is "both an expression of and participation in the discipline" (p. 382).
LaCelle-Peterson, M.W., & Finkelstein, M.J. (2003)	It is dependent upon institutional support.
Dall'Alba, G. (2005), Schwill, S. A. (2008)	Novice teachers should practice actual teaching tasks under the guidance of a mentor or expert teacher.
Pickering, A. M. (2006), Ross, D. (1986)	Training programs can influence novice teachers through the use of best practices.
Leuddeke, G. (2003)	Encourage university professors to become certified teachers.
Parker, J. (2002), Rust, C. (2000)	Reflection on practice is critical.

Differentiating between subject matter expertise and teaching. The last part of this framework is information about how subject matter expertise is conveyed and how that is different from teaching expertise. Most authors who have addressed this seem to explicitly differentiate between the two, yet do not directly address the difference between them.

Table 3

Teaching expertise vs. subject matter expertise

Author	Evidence
Bass, H. (2006)	Only recently has the fact that subject matter expertise is different from, and feeds into, the ability to teach been acknowledged.
Eble, K. (1983)	"In almost all graduate programs, the main emphasis is on mastery of the knowledge of a field and demonstration of research competence in some aspect of that field. These pressures squeeze out all other study and even diminish the attention a candidate might give to such an important matter as developing teaching skills. Moreover, the bias against formal "education" tends to place a barrier between subject matter departments and those engaged in education. It is a rare graduate student who can cross such a barrier to include work in pedagogy or even to seek guidance about teaching and learning. The result is that the practical experience large numbers of graduate students receive as teaching assistants is the chief form of specific preparation for teaching. Paradoxically, Ph.D. candidates in many prestigious institutions and those whose records win them other kinds of support can escape specific preparations for teaching altogether. ...But at best, the acquiring of teaching skill across the disciplines is haphazard and little informed or reinforced by the resources within the university" (p. 124-5).
Richlin, L. (1993a)	As far back as 1949 it was noted that people who enter the professoriate do so with subject matter expertise but without the knowledge of or demonstrated ability in teaching.

Conclusion

If one accepts that graduate study is what prepares one for future employment as part of the professoriate, then one must hope that one's preparation is amply sufficient. Generally, though, graduate programs ignore everything related to the job of faculty member except research (Erickson, 1984; Ross, 1996; Speck, 2003; Ramsden, 2003; et al.). Graduate students have demonstrated knowledge of their subject matter – thorough testing on knowledge of the field is part of every program's degree qualifications. Why would graduate programs completely ignore such a basic qualification as that of preparation to teach? It is a fact that one must have an advanced, post-baccalaureate degree or proven, recognized subject- matter expertise to get a job in the professoriate. It is a further fact that an advanced degree programs at the master's and doctoral level give one in-depth knowledge of the chosen field of study.

Some institutions of higher education have opened centers for teaching, which are called various names. These, by definition, self-select for instructors who are concerned about what they are doing and would be able to uncover the information on how to improve their teaching all on their own. One might challenge that this is too little, because of the self-selection aspect. One might also challenge that it is too late, because these instructors are already teaching, but in fact, it will never be too late because there is always a new crop of students. These centers explicitly offer pedagogical training – helping faculty to build better courses: assignments, grading, teaching techniques, etc. Some highly regarded centers offer pedagogical theory so that the faculty members are grounded in theory for good teaching. The fundamental

problem with these centers for teaching is that 1) it is too little, and 2) they are voluntary, and 3) participation is frequently informal. Faculty can attend or not as they choose.

My argument is that this voluntary process does not work well to improve teaching uniformly across all disciplines at an institution. Since it is a voluntary process, the teachers that participate are driven by something other than remuneration or recognition. They seem to be driven by a desire to be the best teachers they can be (Beatty, 1998). The information provided by these centers is either not being disseminated among or not being actively used by members of the professoriate. If it were, it would be reasonable to expect the average level of teaching would rise, and the average learning level of the students would rise: higher grad rates, higher retention, shorter length of time in the school. There would be some known link between preparation to teach in higher education and student evaluations of teaching as evidence of the success of that preparation.

I believe that this has been ignored for several reasons. First, educational administrators are largely selected from the faculty, and these administrators draw from what they know. Second, higher education is far less regulated at the state and fed levels than public education at the K-12 level, primarily because higher education is optional. However, the law of the marketplace is making a college degree of any kind a necessary part of one's preparation for the workplace. Third, even though the expectations of education on the part of society have changed, education has changed very little. Even in Japan, which invented reciprocal loyalty, higher education is still

the late-medieval institution that it was when it was imported to the US from Europe. Also, it is evident that there is a certain amount of inertia on the part of faculty combined with resistance to change. Finally, there is little, if any, reward tied to being a good teacher. There is a great deal of reward tied to getting published for one's research.

I believe that this issue is a timely one in academia today. In the current atmosphere of student entitlement and service orientation, colleges and universities are rushing to develop faculty teaching abilities through the creation of HRD programs on teaching (Centers for Teaching Excellence and the like on campuses, which have volunteer-based programs in which faculty members may participate). This seems to be a band-aid solution, treating the symptom and not the underlying problem. It is my hope that this study will provide a solid foundation to treat the underlying problem.

Chapter Three: Research Design and Methodology

The purpose of this study was to provide graduate schools, their faculty members, and their students with quantitative evidence that can guide them about the type of preparation necessary to support careers in the professoriate. The major research problem was to determine whether there is a relationship between how faculty members learn to teach and how effective they are at teaching undergraduates, as judged by student evaluations of their courses. If such a relationship exists, it is the first step in establishing that chain of quantitative evidence that could lead to guidance about preparation to teach in a university setting.

This chapter first discusses the research design procedures in terms of the plan of action, including the method and methodology, the population, the data that will be collected and how they will be collected, and the plan for analyzing that data. Next, the concepts of validity and reliability will be explored in relation to the plan of action, followed by a discussion of ethical issues related to the proposed plan. Finally, the strengths and weaknesses of the proposed approach will be identified and discussed. Most supplementary materials, such as an example of the survey instrument, informed consent and release form, and some of the data tables, are included in the appendices.

Research Plan

Method and methodology are terms that tend to be confused when discussing research. For the purposes of this research plan, the methods section will address the specific things that will be done to complete the research. The overall methodology is quantitative, survey-based research, chosen because the proposed research problem

examining relationships statement necessitates a quantitative approach. In other words, the methodology addresses the "what kind of research will be done?" question, while the methods address the specific "how exactly will the research be done?" question.

Methodology and Methods

Not only does the research problem statement necessitate a quantitative approach, the field currently has a gap in quantitative research regarding how teaching skills are addressed for graduate students who plan to pursue careers in the professoriate. In researching this question for the past four years, I have yet to uncover even one published article that includes quantitative research on this topic. There is substantial qualitative research, as outlined in the previous literature review. However, quantitative research on this question does not currently exist. The specific method for this research will consist of an initial, survey-based approach to determine the actual sample, followed by multiple regression analysis to determine the relationship between the independent and dependent variables.

The survey was web-based. It contained one welcome page to secure consent, one page containing the entire survey with a right vertical scroll bar, and a thank-you page that appeared when the survey was submitted. The survey primarily contained closed questions divided into three sections, with an open-ended question requesting additional comments at the end of each of the three sections. It is important to note here that the survey was based on a survey that was developed, reviewed by experts, and pilot-tested during a class at the University of Minnesota during fall semester of 2008. The original survey cannot be used because the primary research question has changed.

However, most of the questions from the survey were retained for use in the new survey proposed by this research plan. More discussion of how the data will be collected appears later in this chapter.

The additional survey questions were designed with social exchange theory in mind. Dillman (2007) wrote that social exchange theory, with respect to survey design and response, necessitates that one pay attention to increasing the rewards for responding, to reducing the perceived costs of responding, and to establishing trust so that the first two of these actually happen. The e-mail asking people to participate, as well as the welcome page on the survey itself, helped address each of these aspects. Related to this theory, I believe that people who are faculty members wanted to participate for one main reason: participating will help to shed light on what preparation for teaching, or lack of it, does to student ratings of teaching (SRTs), which are used in so many institutions of higher education as part of the reward structure.

Target Population

Because the purpose of this research is ultimately to support a change to graduate education, it makes sense to target degree-granting institutions in this project. Therefore, the target population for this study consists of all individuals who are employed as faculty members at a U.S. academic institution designated by the Carnegie Foundation for the Advancement of Teaching as a doctoral-granting institution.

The sample frame used to identify these individuals is the class schedule database for a particular U. S. doctoral-granting institution of higher education. This database contains the names and e-mail addresses for everyone who taught a class for

that institution during the fall semester of 2010, regardless of standing at the institution. This equates to a population of 5,414 unique individuals (N. Kopka, personal communication, December 16, 2010).

The survey will be used as the entry point to select faculty members who taught during the Fall semester of 2010, for two reasons. First, all course instructors are required to distribute Student Ratings of Teaching, and the database thus contains all instructors who were required to distribute SRTs in a given time period. Second, by using one institution, it is guaranteed that the same SRT was used by all course instructors.

Sampling Procedure

A simple random sample was selected from the frame by using a random number generator to select items from a list of names and e-mails generated from the class schedule database. Because neither the content area nor classification of the respondent were being considered, the sample was truly a random sample of individuals across the doctoral-granting institution.

Sample size. The survey sample consists of 359 cases, which will be weighted for survey non-response to become 1,197 cases. The calculations are as follows. The size of the sample should be equal to the z-score squared times the square of the variance, divided by the square of the error. This is more easily written $\frac{z^2 \times s^2}{e^2}$. This is the standard formula used to derive survey sample sizes (Dillman, 2007). For the

purposes of this project, I want to have a 95% confidence level, no more than 5% error, and a 0.50 variance. The math works out in this fashion:

$$\text{Sample size} = \frac{(1.96)^2 \times (0.5)^2}{(0.05)^2}$$

$$= \frac{3.8416 \times 0.25}{0.0025}$$

$$= 359$$

To compensate for response rate, I would assume a pessimistic 30% response rate, which would bring the sample size up to 1,197 ($359 = 0.3(x)$ where x=1197). This pessimistic assumption is based on, first, a meta-analysis of survey response rates that indicated that web-based response rates were 11% lower than non-web-based survey response rates (Manfreda, Bosnjak, Berzelak, Hass, & Vehovar, 2008). Second, a short search for survey research that was web-based and took as its primary population people in academia resulted in articles that had response rates between 41% and 70% (Bulot & Johnson, 2006; Couper & Miller, 2008; Kahanov, Loebsack, Masucci, & Roberts, 2010; Kittleson & Brown, 2005; Menachemi, N., 2011; Mourad, Kohlwes, Maselli, & Auerbach, 2010). If I were to assume that social exchange theory would be at work for my sample, and I could achieve the 70% response rate that social exchange theory regularly receives (Dillman, 2007), my sample size would be 359=0.7x, or 513, rounded up because one cannot have a partial person.

For this sample, the data will be drawn from the SRT database at the institution in question. How many data points are drawn will depend on the survey response.

Data collection techniques. Initial data was collected for this project using Survey Monkey, an online survey tool. The principles of social exchange theory were

employed to collect data. More specifically, I used e-mail to send out a pre-survey notice, and then I sent an e-mail with a link to the survey. I gave respondents approximately two weeks to respond. Because the institution's Institutional Review Board (IRB) is more restrictive than social exchange theory, I was not able to do any follow-up contacts with the members of the sample. Data was input into Survey Monkey automatically as the respondents complete the survey, and the data was stored in Survey Monkey until it was time to extract it for analysis purposes. A copy of the survey questions, including the consent form, may be found in Appendix A. After collecting the survey information, I received the scores for the first four questions of the institution's SRT for those respondents who granted permission for their scores to be collected.

Table 4 will help to illustrate exactly what data I collected, as well as whether the data element is a variable, and what type of variable it will be. More detail on validity and reliability appears in Table 4 below.

Table 4

Variable Information

Variable	Type	Measure	Validity Concerns	Reliability Issues
Teaching preparation	Independent	Categorical – self reported	None	None
Scores on student evaluations	Dependent	Interval and continuous	Maturation – more experienced participants may have better scores	None
Position title	Independent	Categorical	None	None
Experience	Independent (moderator)	Ratio	None	None
Terminal Degree held	Independent (moderator)	Categorical	None	None

Data Analysis Procedures

The sample of 1,277 people employed as instructors was designed to capture 359 full opinion surveys. Most of the questions on the survey were categorical in nature. The survey acted as the entry point to the database that contains the SRT data, and said data is both interval and continuous. Multiple regression was used to analyze this data in relation to teaching preparation. All of these tests were generalizable across the population, because the sample is simple and random. Further, the critical survey questions did not deal with institution type or position title, so this data was not critical to my data collection and therefore my analysis. Instead, this type of analysis may shed light on confounding variables after the survey data has been collected.

Missing data and nonresponse error. Surveys that do not include responses to the critical questions were excluded from the final sample. These questions provide specific permission to access the SRT scores, and provide information about the number of years the respondent has taught at any level, as well as the number of semesters the respondent has taught at the particular institution at which the respondent would have SRT scores. If the respondent did not have SRT scores available in the institution's database, then that survey was also excluded from the sample frame. It may be possible to perform post-hoc analysis of the excluded surveys at a later time, but such an analysis is outside the scope of this research.

In order to minimize the impact of unit nonresponse as a source of error, data will be adjusted for nonresponse through the use of statistical weighting techniques (Dillman, 2007). Specifically, the survey was weighted for a 30% response rate. Regarding item nonresponse, as stated above, any cases missing the critical complete items will be considered as survey nonresponse. Critical complete items include items 1, 11 and 14. Because there are only three critical complete items, and because one of these items is the specific permission to access SRT scores, and the other items record information about the respondent's preparation for and number of years teaching, it is not feasible to identify item non-response and use a lower N. Item nonresponse on either of these items must be considered survey nonresponse.

Reliability. The notion of reliability with respect to survey design is an important one. Reliability is the consistency or dependability of a score. If a survey is reliable, then if one takes a survey multiple times, the same results would be received

from the same person. In other words, the trait being measured is thought of as being stable over time (Gall, Gall, & Borg, 2007). Reliability in the case of this research is generally a non-issue. The questions asked for information regarding past training, which is a stable item in a faculty member's past, and cannot change. Thus, if the same faculty member were to take the survey multiple times, the responses will always be the same.

However, there may be a question of reliability in that respondents may have inflated or deflated the number of years of teaching, since two of the questions, regarding number of years and semesters the respondent has taught, are self-reported. It is possible that respondents may not have remembered how long they have taught at the specific institution, especially if they have semester-to-semester contracts. However, these particular survey items address moderator variables, not key variables, and so may play only a supporting role in the analysis. Further, in this case, social exchange theory comes into play. Since the survey responses will be confidential, and since each respondent will be assured of this fact in the initial introductory e-mail, the invitation-to-participate e-mail, and in the welcome screen for the survey, the risk for providing inflated or deflated scores is diminished, making it more likely that the respondents will answer honestly.

Reliability for the SRT is also a concern. While the overall concept of reliability with respect to SRTs was discussed above on page 34, it is important to note that the documentation for the development of the SRT at this particular institution makes use of the same research, the same authors, and a rigorous assessment method in the

development process. In the interest of preserving confidentiality, that documentation is not included in the references section of this paper. While the institution at which the SRTs were used does not provide any reliability data, the internal reliability of the data being used in the study can be reported upon. Please see the next chapter for more information.

Validity. Instrument validity is also an important element of survey design. Instrument validity says that the survey measures what it is designed to measure, does not measure the things it is not designed to measure, and allows one to interpret the responses accurately (Gall, Gall, & Borg, 2007). Instrument validity of the survey was determined by subject matter expert review, as well as pilot testing. Instrument validity information of the SRT is, unfortunately, not published by the institution that uses the SRT.

Some of the threats to the internal validity of this research design include nonresponse for survey-based research, and type 1 errors because of data collection problems. I minimized non-response error by using social exchange theory to develop and implement surveys. Social exchange theory indicates that the likelihood of receiving a response to a questionnaire is higher when the respondent expects that the rewards of responding outweigh the costs (Dillman, 2007). What this means in practical terms is that the survey needed to appear important, needed to be interesting, easy, and short, and the survey process had to show positive regard for the respondents and gratitude for their responses. I avoided type 1 errors by setting my alpha value lower. There is one limitation to the external validity of the findings. This limitation,

coverage error, was dealt with by randomly sampling without replacement from the entire population of faculty members who taught during the fall semester of 2010. This ensured that all members of the survey population had an equal chance of being selected for the survey.

Other measures. In survey design, there are four additional, important measures of survey quality beyond validity and reliability, but which contribute to these. These are sampling variability; coverage; nonresponse, which was discussed above; and measurement. Regarding sampling variability, the sample size was sufficiently large that estimates based on the total sample should be subject to no more than moderate sampling error. However, sampling error can be quite substantial in estimating the characteristics of small subgroups of the population. Estimates of the sampling errors associated with various measures are included in the methodology report for the survey and in the basic publications. Coverage for the population is believed to be excellent, as the population and the sample frame are equivalent. Several of the key variables in this survey were difficult to measure and thus were relatively prone to measurement error. In order to reduce measurement error, the instrument was pretested, using talk-throughs and a pilot. In addition, the survey instrument was reviewed by experts. Minor wording changes resulted from these processes.

Ethics and Human Relations

My research poses a threat to human relations only if the participants are over-surveyed at their institutions. Otherwise, using the social exchange theory concepts mentioned above to encourage participation will minimize any potential threat in this

area. Also, it will be easy for participants to respond, as there will be no personal contact. Confidentiality will be maintained because no information will be reported that will make it possible to identify an individual subject. Hard copies of data will be stored in a locked cabinet for six years after the conclusion of the project. In these ways, the initial perceived ethical and human relations issues will be addressed. If further ethical or human relations issues arise during the time the project is proceeding, they will be addressed as deemed appropriate by the institution's IRB guidelines.

Strengths and Weaknesses

This proposed research intends to fill an existing gap in the field regarding quantitative data on how faculty members learn to teach; in this respect, the approach is, itself, a strength. Another strength was uncovered in completing the research on teaching effectiveness as judged by SRTs. While I knew that SRTs were controversial as a way of evaluating teaching, I had not realized that the link between SRTs and preparation to teach had not been explored in higher education. I uncovered no published research on this link, nor was it mentioned, directly or indirectly, in the research I did find.

There are several weaknesses within the proposed research design. One of these was identified above: if faculty members are over-surveyed at their institution, they may be less-inclined to participate. However, in a brief, informal discussion with colleagues at this institution, this weakness seems to be less critical than originally thought. Surveys seem to be received very infrequently by faculty members at the institution that will be contacted as part of this research. Another weakness is within the survey

design itself. The specific permission to access SRT scores since 2009 will require a name and e-mail address. This could make confidentiality a question. The way to guard against this is to convert names and e-mails to identifier codes and then destroy the conversion table once the SRT scores are retrieved. A final weakness is the sheer volume of e-mails that must be sent in order to assure a 30% response rate. This is a risk that will have to be absorbed, rather than ameliorated or avoided.

Chapter Four: Results of the Study

The purpose of this study is to determine if there is a relationship between how faculty members learn to teach and how effective they are at teaching undergraduates, as judged by student evaluations of their courses. This chapter will present the results of the statistical analysis.

The data for the dependent variables was interval data: the mean of the class set of SRTs that were turned in at the end of Fall Semester, 2010. More accurately, each of the first four questions of the SRT had a class set of responses, and what was used was not the hard data from the responses, but instead the mean of those responses. Each class set was treated as a separate data point; no scores were averaged, even if one person had more than one class set. The total number of usable responses (N = 176) resulted in a 14.7% response rate for the survey. The data for the independent variables was taken from the survey that served as the entry point into the SRT database. The survey data pertinent to the research questions was converted to dummy variables to make them dichotomous. For questions where the answer on the survey questions pertinent to the research questions was either "None" or "No", the data was coded "1". For questions where the answer on those survey questions was anything other than "None" or "No", the data was coded "0". Table 31 in Appendix G shows the questions mapped to the specific independent variables, and all of them were used in the final analysis. Next, the data was imported into SPSS, a standard statistical analysis software package. Multiple regression analysis was run, using backward entry on the

independent variables. Backward entry allows the software package to find the best fit of the independent variables into the regression equation (Field, 2005).

The four questions from the SRT that are used as dependent variables are found in Table 5 below.

Table 5

SRT Question Text

SRT Question Number	Question Text
1	The instructor was well prepared for class.
2	The instructor presented the subject matter clearly.
3	The instructor provided feedback intended to improve my course performance.
4	The instructor treated me with respect.

I ran full descriptive statistics for each question, and the means and standard deviations are shown in Table 18 in Appendix B. A short summary table for each question appears in Table 6 below.

Table 6

Descriptive Statistics for SRT Questions

Question	N	Minimum	Maximum	Mean	Std. Deviation
InstrWellPrep-Mean (Q1)	176	4.24	6.00	5.5488	.39026
InstrPresClear-Mean (Q2)	176	2.88	6.00	5.2540	.57529
InstrProvFeedback-Mean (Q3)	176	3.12	6.00	5.2056	.56195
InstrRespect-Mean (Q4)	176	4.54	6.00	5.6374	.29364
Valid N (listwise)	176				

Note. 1=strongly disagree, 2=disagree, 3=somewhat disagree, 4=somewhat agree, 5=agree, 6=strongly agree.

I ran full descriptive statistics for each predictor variable as well. The definitions of each variable are shown in Table 7 below, and the descriptive statistics are shown in Table 8.

Table 7

Predictor Variables Included in Multiple Regression Analysis

SPSS Variable	Long Description
TrainingRcvd_ConductClass	How much training did you receive (if any) on how to conduct classes?
TrainingRcvd_ConstrSyll	How much training did you receive (if any) on how to construct a syllabus?
TrainingRcvd_Grading	How much training did you receive (if any) on grading student work?
TrainingRcvd_GiveFeedback	How much training did you receive (if any) on giving feedback to students?
TrainingRcvd_ChooseInstrMeth	How much training did you receive (if any) on choosing instructional methods?
YearsTaughtThisInstitution	How many years have you taught at any level?
TaughtPreschool	How many years have you taught at the preschool level?
TaughtK-12	How many years have you taught at the K-12 level?
TaughtUndergrad	How many years have you taught at the undergraduate level?
TaughtGrad	How many years have you taught at the graduate level?
TaughtBusIndust	How many years have you taught in business or industry?
YearsTaughtThisInstitution	How many years have you taught at this institution?

Table 8

Descriptive Statistics for Predictor Variables

Variable	Minimum (N)	Maximum (N)	Mean	Std. Deviation
TrainingRcvd_ConductClass	0 (73)	1 (103)	.41	.494
TrainingRcvd_ConstrSyll	0 (102)	1 (74)	.58	.495
TrainingRcvd_Grading	0 (64)	1 (112)	.36	.482
TrainingRcvd_GiveFeedback	0 (73)	1 (103)	.41	.494
TrainingRcvd_ChooseInstrMeth	0 (89)	1 (87)	.51	.501
YearsTaughtAnyLevel	1 (a)	5 (a)	3.85	1.197
TaughtPreschool	0 (169)	1 (7)	.04	.196
TaughtK-12	0 (132)	1 (44)	.25	.434
TaughtUndergrad	0 (12)	1 (164)	.93	.253
TaughtGrad	0 (92)	1 (84)	.48	.501
TaughtBusIndust	0 (134)	1 (42)	.24	.427
YearsTaughtThisInstitution	1 (b)	5 (b)	3.43	1.272
Valid N (listwise)	176	176		

Note. a. 1=less than 1 year (N=13). 2=1 year (N=8). 3=1.5 to 5 years (N=39). 4=5.5 to 10 years (N=49). 5=more than 10 years (N=67)
b. 1=less than 1 year (N=15). 2=1 year (N=24). 3=1.5 to 5 years (N=58). 4=5.5 to 10 years (N=28). 5=more than 10 years (N=51)

Also in Appendix B, Table 19 shows the Pearson correlations for each variable of interest in the main effect model, aggregated by question. I also examined the casewise diagnostic statistics for extreme cases in the data, and the information for this examination is included in Appendices C through F. Finally, I examined the descriptive statistics for violations of the assumptions of multiple regression, and the information for this examination is contained in Table 9 below.

Table 9

Significant Variables By Question

SRT Question Number	Predictor Variable	Beta	*p*
1	YearsTaughtThisInstitution	.100	<.001
1	TrainingRcvd_ChooseInstrMeth	-.212	<.001
2	TrainingRcvd_ConstrSyll	-.158	.092
2	TrainingRcvd_ChooseInstrMeth	-.212	.023
2	YearsTaughtThisInstitution	.141	<.001
3	TrainingRcvd_ChooseInstrMeth	-.242	.005
3	TaughtK-12	.205	.037
4	TrainingRcvd_ChooseInstrMeth	-.107	.038
4	TrainingRcvd_Grading	-.119	.097
4	TrainingRcvd_ChooseInstrMeth	-.107	.072
4	TrainingRcvd_GiveFeedback	.170	.029

SRT Question One

In looking at the first question on the Student Ratings of Teaching form (see Table 5), students are asked to agree or disagree with the statement that the instructor was prepared for class. A scale of with one, equating to "Strongly Disagree," through six, equating to "Strongly Agree," is used. In using the backwards entry method in SPSS, the following equation resulted:

Q1Score = 5.311+(0.100YearsTaughtThisInstitution)+(-0.212TrainingRcvd_ChooseInstrMeth).

This equation explains 16.1% of the variation in the dependent variables ($F = 16.067$, $p < .001$).

Assumptions. Multicollinearity, which is a strong correlation between at least two predictors in the above model, poses a threat to the validity of the above analysis. Regarding multicollinearity, the main index in SPSS that warns of this problem is the Variance Inflation Factor (VIF), and for question one the score average is 2.3, and all tolerance statistics are above 0.2. According to Field (2005), these scores indicate that we can conclude that collinearity is not a problem for this model. The assumption of homoscedasticity, which is the assumption that the residuals at each level of the predictor variable have similar variances (Field, 2005), is met: the graph of the standardized residuals is a slightly curved line. The assumption of linearity is also not a problem because everything on the scatterplots of the standardized residuals against the standardized predicted values is roughly linear. The assumption of independent, random, and normally distributed errors is met. The Durbin-Watson statistic is 2.032, which is neither less than 1 nor more than 3, and indicates that the residuals in the model are independent. Please view Appendix C, Table 20 and Table 21, with accompanying text, for an explanation of the casewise diagnostics for SRT question one.

Model summary data. The SPSS tool stepped through each of the independent variables from the survey data and produced 10 below for question one.

Table 10

Model Summary Data for Question One

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.442[a]	.195	.136	.36277	.195	3.294	12	163	.000	
2	.442[b]	.195	.141	.36167	.000	.009	1	163	.926	
3	.442[c]	.195	.146	.36062	.000	.040	1	164	.842	
4	.441[d]	.195	.151	.35958	.000	.044	1	165	.835	
5	.440[e]	.194	.155	.35872	-.001	.209	1	166	.648	
6	.439[f]	.193	.159	.35787	.000	.198	1	167	.657	
7	.437[g]	.191	.162	.35724	-.002	.405	1	168	.525	
8	.431[h]	.186	.162	.35726	-.005	1.022	1	169	.314	
9	.426[i]	.182	.162	.35717	-.004	.912	1	170	.341	
10	.416[j]	.173	.159	.35789	-.008	1.698	1	171	.194	
11	.401[k]	.161	.151	.35951	-.012	2.571	1	172	.111	2.032

Note. a. Predictors: (Constant), YearsTaughtThisInstitution, TrainingRcvd_ConstrSyll, TaughtUndergrad, TaughtPreschool, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
b. Predictors: (Constant), YearsTaughtThisInstitution, TrainingRcvd_ConstrSyll, TaughtUndergrad, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
c. Predictors: (Constant), YearsTaughtThisInstitution, TrainingRcvd_ConstrSyll, TaughtUndergrad, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
d. Predictors: (Constant), YearsTaughtThisInstitution, TrainingRcvd_ConstrSyll, TaughtUndergrad, TaughtK-12, TaughtBusIndust, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
e. Predictors: (Constant), YearsTaughtThisInstitution, TrainingRcvd_ConstrSyll, TaughtUndergrad, TaughtK-12, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
f. Predictors: (Constant), YearsTaughtThisInstitution, TrainingRcvd_ConstrSyll, TaughtUndergrad, TaughtK-12, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
g. Predictors: (Constant), YearsTaughtThisInstitution, TrainingRcvd_ConstrSyll, TaughtUndergrad, TaughtK-12, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_GiveFeedback
h. Predictors: (Constant), YearsTaughtThisInstitution, TaughtUndergrad, TaughtK-12, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_GiveFeedback
i. Predictors: (Constant), YearsTaughtThisInstitution, TaughtUndergrad, TaughtK-12, TrainingRcvd_ChooseInstrMeth
j. Predictors: (Constant), YearsTaughtThisInstitution, TaughtK-12, TrainingRcvd_ChooseInstrMeth
k. Predictors: (Constant), YearsTaughtThisInstitution, TrainingRcvd_ChooseInstrMeth
l. Dependent Variable: InstrWellPrep-Mean

SRT Question Two

The second SRT question asks students to evaluate an instructor's clear presentation of information. It uses the same six-point scale, from Strongly Disagree at one to Strongly Agree at six. Again using backwards entry in SPSS, the following equation resulted:

Q2Score = 4.910+(-0.158TrainingRcvd_ConstrSyll)+ (-0.212TrainingRcvd_ChooseInstrMeth) +(0.141YearsTaughtAnyLevel). This equation explains 16% of the variation in the dependent variables ($F = 10.881$, $p < .001$).

Assumptions. Regarding multicollinearity, VIF score average is 1.22, and all tolerance statistics are above 0.2. The assumption of multicollinearity has not been violated. The assumption of homoscedasticity is met – the graph is a slightly curved line. The assumption of linearity is met because everything on the scatterplots is roughly linear. The assumption of independent, random, and normally distributed errors is met. The Durbin-Watson statistic is 2.087, which is neither less than 1 nor more than 3. Please view Appendix D, Table 23 and Table 24, with accompanying text, for an explanation of the casewise diagnostics for SRT question two.

Model summary data. The SPSS tool stepped through each of the independent variables from the survey data and produced Table 11 below for question two.

Table 11

Model Summary Data for Question Two

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.423[a]	.179	.118	.54027	.179	2.952	12	163	.001	
2	.422[b]	.178	.123	.53864	.000	.011	1	163	.916	
3	.422[c]	.178	.129	.53703	.000	.016	1	164	.898	
4	.421[d]	.178	.133	.53568	.000	.165	1	165	.685	
5	.420[e]	.176	.137	.53452	-.001	.280	1	166	.597	
6	.417[f]	.174	.139	.53374	-.002	.506	1	167	.478	
7	.415[g]	.172	.143	.53269	-.002	.337	1	168	.563	
8	.411[h]	.169	.144	.53214	-.003	.652	1	169	.420	
9	.406[i]	.164	.145	.53198	-.004	.893	1	170	.346	
10	.399[j]	.160	.145	.53200	-.005	1.013	1	171	.316	2.087

Note. a. Predictors: (Constant), YearsTaughtThisInstitution, TrainingRcvd_ConstrSyll, TaughtUndergrad, TaughtPreschool, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
b. Predictors: (Constant), YearsTaughtThisInstitution, TrainingRcvd_ConstrSyll, TaughtUndergrad, TaughtPreschool, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
c. Predictors: (Constant), YearsTaughtThisInstitution, TrainingRcvd_ConstrSyll, TaughtPreschool, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
d. Predictors: (Constant), YearsTaughtThisInstitution, TrainingRcvd_ConstrSyll, TaughtPreschool, TaughtK-12, TaughtBusIndust, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
e. Predictors: (Constant), YearsTaughtThisInstitution, TrainingRcvd_ConstrSyll, TaughtPreschool, TaughtK-12, TaughtBusIndust, TaughtGrad, TrainingRcvd_ChooseInstrMeth, YearsTaughtAnyLevel
f. Predictors: (Constant), TrainingRcvd_ConstrSyll, TaughtPreschool, TaughtK-12, TaughtBusIndust, TaughtGrad, TrainingRcvd_ChooseInstrMeth, YearsTaughtAnyLevel
g. Predictors: (Constant), TrainingRcvd_ConstrSyll, TaughtPreschool, TaughtBusIndust, TaughtGrad, TrainingRcvd_ChooseInstrMeth, YearsTaughtAnyLevel
h. Predictors: (Constant), TrainingRcvd_ConstrSyll, TaughtPreschool, TaughtBusIndust, TrainingRcvd_ChooseInstrMeth, YearsTaughtAnyLevel
i. Predictors: (Constant), TrainingRcvd_ConstrSyll, TaughtPreschool, TrainingRcvd_ChooseInstrMeth, YearsTaughtAnyLevel
j. Predictors: (Constant), TrainingRcvd_ConstrSyll, TrainingRcvd_ChooseInstrMeth, YearsTaughtAnyLevel
k. Dependent Variable: InstrPresClear-Mean

SRT Question Three

The third SRT question asks students to evaluate the feedback an instructor gave. It uses the same six-point scale, from Strongly Disagree at one to Strongly Agree at six. Again using backwards entry in SPSS, the following equation resulted: Q3Score = 5.277+ (-0.242TrainingRcvd_ChooseInstrMeth)+(0.205TaughtK-12). This equation explains 9.0% of the variation in the dependent variables ($F = 8.566$, $p < .001$).

Assumptions. The VIF score average is 1.078, and all tolerance statistics are above 0.2, so the assumption of multicollinearity has not been violated. The assumptions of homoscedasticity and linearity are met because the graph is roughly linear and slightly curved. The assumption of independent, random, and normally distributed errors is met, because the Durbin-Watson statistic is 1.954, which is neither less than 1 nor more than 3. Please view Appendix E, Table 26 and Table 27, with accompanying text, for an explanation of the casewise diagnostics for SRT question three.

Model summary data. The SPSS tool stepped through each of the independent variables from the survey data and produced Table 12 below for question three.

Table 12

Model Summary Data for Question Three

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.351[a]	.123	.059	.54524	.123	1.908	12	163	.037	
2	.350[b]	.123	.064	.54368	.000	.064	1	163	.801	
3	.349[c]	.122	.069	.54226	.000	.139	1	164	.710	
4	.348[d]	.121	.074	.54090	.000	.168	1	165	.682	
5	.346[e]	.120	.078	.53964	-.001	.219	1	166	.641	
6	.342[f]	.117	.080	.53890	-.003	.543	1	167	.462	
7	.337[g]	.114	.082	.53836	-.003	.663	1	168	.417	
8	.334[h]	.111	.085	.53747	-.002	.434	1	169	.511	
9	.325[i]	.105	.084	.53772	-.006	1.161	1	170	.283	
10	.309[j]	.096	.080	.53903	-.010	1.837	1	171	.177	
11	.300[k]	.090	.080	.53913	-.006	1.065	1	172	.304	1.954

Note. a. Predictors: (Constant), YearsTaughtThisInstitution, TrainingRcvd_ConstrSyll, TaughtUndergrad, TaughtPreschool, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
b. Predictors: (Constant), TrainingRcvd_ConstrSyll, TaughtUndergrad, TaughtPreschool, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
c. Predictors: (Constant), TrainingRcvd_ConstrSyll, TaughtPreschool, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
d. Predictors: (Constant), TrainingRcvd_ConstrSyll, TaughtPreschool, TaughtK-12, TrainingRcvd_Grading, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
e. Predictors: (Constant), TrainingRcvd_ConstrSyll, TaughtPreschool, TaughtK-12, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
f. Predictors: (Constant), TaughtPreschool, TaughtK-12, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
g. Predictors: (Constant), TaughtK-12, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
h. Predictors: (Constant), TaughtK-12, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, YearsTaughtAnyLevel
i. Predictors: (Constant), TaughtK-12, TaughtGrad, TrainingRcvd_ChooseInstrMeth, YearsTaughtAnyLevel
j. Predictors: (Constant), TaughtK-12, TrainingRcvd_ChooseInstrMeth, YearsTaughtAnyLevel
k. Predictors: (Constant), TaughtK-12, TrainingRcvd_ChooseInstrMeth
l. Dependent Variable: InstrProvFeedback-Mean

SRT Question Four

Question four on the Student Rating of Teaching reads as follows: The instructor treated me with respect. This question asks students to evaluate the feedback an instructor gave. It uses the same six-point scale, from Strongly Disagree at one to Strongly Agree at six. Again using backwards entry in SPSS, the following equation resulted: Q4Score = 5.726+(-0.107TrainingRcvd_ConstrSyll) +(-0.119TrainingRcvd_Grading) +(0.170TrainingRcvd_GiveFeedback)+ (-0.107TrainingRcvd_ChooseInstrMeth). This equation explains 8.0% of the variation in the dependent variables ($F = 3.846$, $p = .005$).

Assumptions. Regarding multicollinearity, VIF score average is 2.2255, and all tolerance statistics are above 0.2, so this assumption is not violated. The assumption of homoscedasticity is met: the graph is a slightly curved line. The assumption of linearity is not violated because everything on the scatterplots is roughly linear. The assumption of independent, random, and normally distributed errors is met because the Durbin-Watson statistic is 2.028, which is neither less than 1 nor more than 3. Please view Appendix F, Table 29 and Table 30, with accompanying text, for an explanation of the casewise diagnostics for SRT question four.

Model summary data. The SPSS tool stepped through each of the independent variables from the survey data and produced Table 13 below for question four.

Table 13

Model Summary Data for Question Four

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.328[a]	.107	.042	.28744	.107	1.635	12	163	.086	
2	.328[b]	.107	.048	.28656	.000	.002	1	163	.963	
3	.327[c]	.107	.053	.28575	.000	.069	1	164	.793	
4	.327[d]	.107	.058	.28497	.000	.086	1	165	.769	
5	.326[e]	.106	.063	.28419	.000	.091	1	166	.763	
6	.324[f]	.105	.067	.28356	-.001	.257	1	167	.613	
7	.314[g]	.099	.067	.28365	-.006	1.109	1	168	.294	
8	.303[h]	.092	.065	.28391	-.007	1.304	1	169	.255	
9	.287[i]	.083	.061	.28453	-.009	1.747	1	170	.188	2.028

Note. a. Predictors: (Constant), YearsTaughtThisInstitution, TrainingRcvd_ConstrSyll, TaughtUndergrad, TaughtPreschool, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
b. Predictors: (Constant), TrainingRcvd_ConstrSyll, TaughtUndergrad, TaughtPreschool, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TaughtGrad, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
c. Predictors: (Constant), TrainingRcvd_ConstrSyll, TaughtUndergrad, TaughtPreschool, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
d. Predictors: (Constant), TrainingRcvd_ConstrSyll, TaughtPreschool, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback, YearsTaughtAnyLevel
e. Predictors: (Constant), TrainingRcvd_ConstrSyll, TaughtPreschool, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback
f. Predictors: (Constant), TrainingRcvd_ConstrSyll, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_ConductClass, TrainingRcvd_GiveFeedback
g. Predictors: (Constant), TrainingRcvd_ConstrSyll, TaughtK-12, TaughtBusIndust, TrainingRcvd_Grading, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_GiveFeedback
h. Predictors: (Constant), TrainingRcvd_ConstrSyll, TaughtBusIndust, TrainingRcvd_Grading, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_GiveFeedback
i. Predictors: (Constant), TrainingRcvd_ConstrSyll, TrainingRcvd_Grading, TrainingRcvd_ChooseInstrMeth, TrainingRcvd_GiveFeedback
j. Dependent Variable: InstrRespect-Mean

Table 14

Regression Equation Summary Data

Question	Regression Constant	Years Taught This Institution	Choosing Instructional Method	Constructing a Syllabus	Years Taught at Any Level	Taught K-12	Grading	Giving Feedback	R	R^2
1	5.311	0.100	-0.212						0.401	0.161
2	4.910		-0.212	-0.158	0.141				0.399	0.160
3	5.277		-0.242			0.205			0.300	0.090
4	5.726		-0.107	-0.107			-0.119	0.170	0.287	0.083

Chapter Five: Discussion, Conclusions, and Implications

The main purpose of this research is to discover if there is a relationship between instructor preparation to teach and scores on Student Ratings of Teaching. In the previous chapter, I described the use of SPSS to conduct a multiple regression analysis and provided the basic results of that analysis. In this chapter, I will present a discussion of those results, as well as their implications. Finally, I will suggest some future direction for research on this topic, and I will provide specific research questions that still exist related to this research.

Regression Equations

The regression equations for each question, disappointingly, do not explain much of what is going on with the dependent variable. The independent variables for each question all explained less than 20% of the change in the SRT scores, so it quickly became clear that there are other things having an effect on the scores. However, finding that there is a weak and positive relationship between preparation to teach and SRT scores is only the first step. Let us now move into specifics for each question.

Question one on the SRT evaluates the instructor's preparedness for class. The independent variables for question one are only accounting for 16.1% of the change in scores, so there is some other effect on the scores than the instructor having had training on instructional methods and having taught for a number of years at the institution under consideration. These variables had the strongest effect on the associated question of all the questions and variables that were part of this research, but still did not account for 83.9% of the change in the scores. Interestingly, the regression constant for the

statement “The instructor was well prepared for class” is 5.31 on a scale of one to six, with six being “Strongly Agree.” It is important to note that if an instructor has never had any training on instructional methods, the score will drop by 0.21 to 5.10. For having taught previously at the institution where the class is taking place, though, the instructor can expect an increase of 0.10 to the score. Thus, the relationship between the score on SRT question one and the independent variables is positive, but weak. It is statistically significant, but practically speaking, not very useful.

Question two on the SRT evaluates whether the instructor presents the content clearly. The independent variables for question two are only accounting for 16% of the change in scores, so there is some other effect on the scores than the instructor having had training on constructing syllabi and on instructional methods, and having taught for a number of years. The relationship, therefore, is positive, weak, statistically significant, but practically not very significant. The base score for the statement “The instructor presented the subject matter clearly” is 4.91 on a scale of one to six, with six being “Strongly Agree.” This was the lowest of the scores among the four questions. If an instructor has never had any training on constructing syllabi, this score will drop by 0.16 to 4.75. If an instructor has never had any training on instructional methods, the score will drop by 0.21 to 4.70. For having taught previously at the institution where the class is taking place, though, the instructor can expect an increase of 0.14 to the score.

The independent variables for question three, which evaluates instructor feedback, are only accounting for 9% of the change in scores. Again, there is some

other effect on the scores than the instructor having had training on constructing syllabi and on instructional methods, and having taught for a number of years. The statement "The instructor provided feedback intended to improve my course performance" has a base score of 5.28 on the same scale as the other questions. If an instructor has never had any training on instructional methods, the score will drop by 0.24 to 5.04. For having taught previously at the institution where the class is taking place, though, the instructor can expect an increase of 0.21 to the score. Again, the relationship is statistically significant, positive, and weak, and is practically not very significant.

For question four on the SRT, the independent variables are only accounting for 8% of the change in scores, so there is some other effect on the scores than the instructor having had training on constructing syllabi, on choosing instructional methods, on giving feedback, and on grading. Interestingly, the regression coefficient for the statement "The instructor treated me with respect" is 5.73 on the same scale of one to six, with six being "Strongly Agree." It is important to note that if an instructor has never had any training on how to construct a syllabus, the score will drop by 0.11 to 5.62. If an instructor has never had any training on grading techniques or practices, the score will drop by 0.12 to 5.61. If an instructor has never had any training on giving feedback to students, the score will rise by 0.17 to 5.90. If an instructor has never had any training on instructional methods, the score will drop by 0.11 to 5.62. All of these scores are statistically significant, but in practice, not at all significant.

Skewedness

In exploring the data, both individual data points and the data set as a whole, I noticed something unusual. Each of these questions has an unexpectedly high mean. While it is somewhat inaccurate to take a mean of means without the N to weight the existing means, the N for each class did not exist, and so I was forced to use a simple mean. Regarding the SRT question means, one would expect with a random sample to have results that fit a bell curve. That is not the case for these scores. Further exploration of this is merited.

Sample Size

The number of data points was initially a cause for concern. Survey research demands a certain response rate, and the response rate for the survey that served as an entry point for this research was 14.7%. This produced only a fraction (49%) of the desired 359 cases. Still, this research did not intend to use the survey results in survey analysis, but to use the results to perform multiple regression analysis against an existing database. Thus, the number of data points was amply sufficient for the purpose.

Database Skewedness

The first step in further exploration is looking at the database that the SRT scores came from. The administrators of the SRT database publish the N and averages for 2008. The data in the database has not changed in any way that is statistically significant since 2008, and so the administrators have not published new information

(D. Glover, personal communication, August 17, 2011). The numbers from 2008 appear in Table 15.

Table 15

Full SRT Database Average Scores, 2008

	Strongly Disagree	Disagree	Somewhat Disagree	Somewhat Agree	Agree	StronglyAgree	Mean
Q1	460	1,049	1,764	8,458	39,826	67,780	5.43
Q2	1,262	2,560	4,841	16,250	39,728	54,601	5.13
Q3	1,615	3,497	7,146	20,576	36,751	49,262	4.98
Q4	687	952	1,921	6,950	33,673	74,648	5.49

Figure 1. Graph of SRT Database Average Scores, 2008

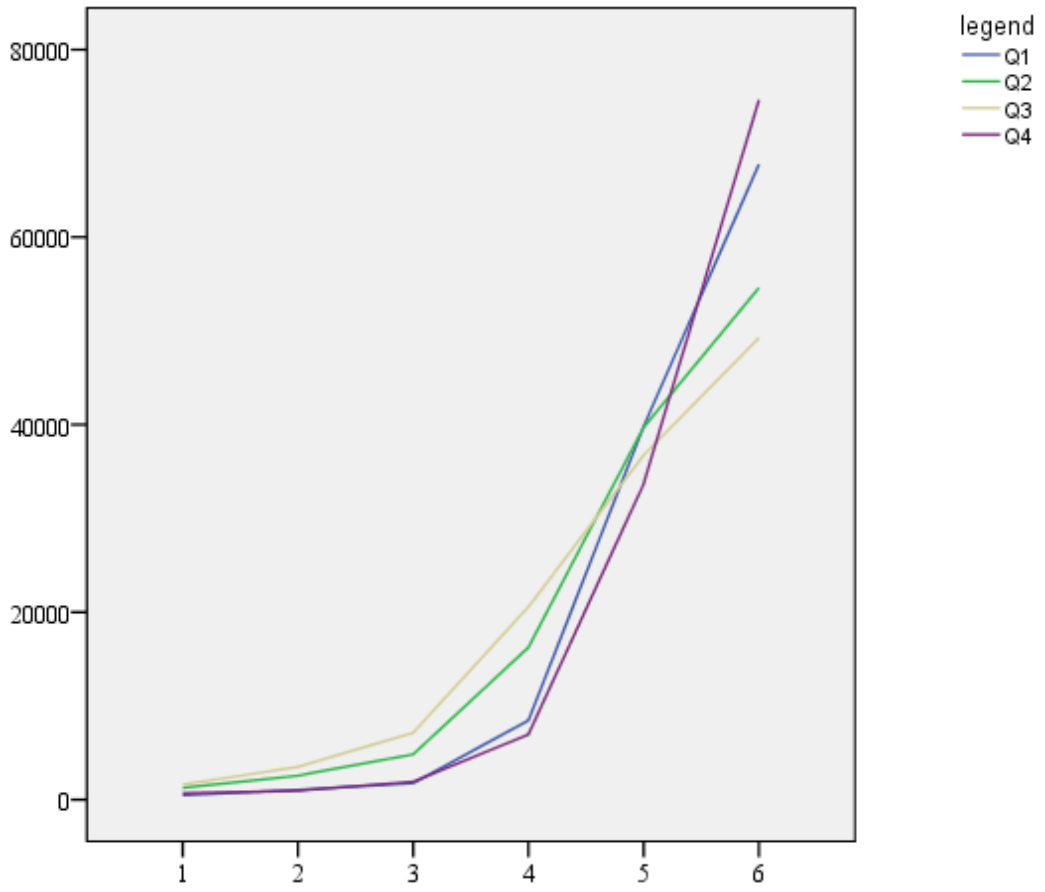

Figure 1. Graphic representation of the data in Table 15, which shows the skewedness of the full SRT database.

The database appears strongly skewed to the left. One would expect that SRT scores over the entire population of classes at a given university would have a more normal distribution. This data seems to have a wedge shape, not a bell curve shape. The data in the sample is likewise skewed to the left when compared with a normal bell curve, but is yet a normal sample when compared to the database from which the sample was drawn. What this means, however, is that additional information is needed

to meaningfully interpret the data. The designers of the SRT for the institution in question relied heavily on Arreola's work (2007) during the design process. Thus, it makes sense to adopt Arreola's decile interpretation of student ratings (p. 121). To do this, the data were divided into 10 equal portions. Each portion was then "defined as representing a 10% interval of rated courses with a defined minimum and maximum course mean" (p. 121). The lowest 10%, 20%, and 30% were defined as Very Poor. Any mean falling in the 40% or 50% interval was defined as Poor, and those in the 60% and 70% were defined as Good. The remaining intervals were defined as Very Good. The ten intervals, as applied to the data in Table 15, result in the following definitions:

Table 16

Decile Interpretation of Student Ratings

Rating	Score	Interpretation
Very Poor	1 to 3.29	"unacceptably low quality insofar as the elements being measured by the student rating form are concerned" (Arreola, 2007, p. 121)
Poor	3.30 to 5.49	"being of low quality and indicates that improvement is needed" (Arreola, 2007, p. 121)
Good	5.50 to 5.69	"being of solid professional quality, and although improvement is always desirable, no improvement is required" (Arreola, 2007, p. 121)
Very Good	5.70 to 6	"high quality…surpassing standard professional performance" (Arreola, 2007, p. 122)

This interpretation adds a level of interest to the results. It becomes clear that the database as a whole needs to be interpreted as having average scores in the "poor" range. Additionally, for this institution, either the instructors are actually good, or there is an upward bias to rate instructors high. Finally, the SRTs do not seem to discriminate well between instructors.

Sample Skewedness

The interpretation of the sample regarding the first four questions also merits new interpretation. No longer does it seem that the base score for each question is as high as previously thought. Instead, it becomes clear that only the last question starts with a base score that could be considered "very good", at 5.71. The rest start with base scores in the "poor" range. Interestingly, the minimum/maximum for the dependent variables were as follows in Table 17:

Table 17

Minimums and Maximums of SRT Questions 1-4

SRT Question	N	Minimum	Maximum
InstrWellPrep-Mean	176	4.24	6.00
InstrPresClear-Mean	176	2.88	6.00
InstrProvFeedback-Mean	176	3.12	6.00
InstrRespect-Mean	176	4.54	6.00
Valid N (listwise)	176		

This information also demonstrates that the sample is strongly skewed to the left, but is still representative of the data in the full database.

Still, this does not necessarily mean that the regression equations are invalid. Rather, when taking into account the skewedness of the database, along with the fact that less than 20% of the change in the ratings is explained by the variables in question, the predictive validity of the equations is questionable.

Further Questions

What other things are affecting the scores? I initially suspected that students do not want to grade instructors poorly, and so start with a 6 in mind, reducing scores from there for real or perceived "infractions". I also wondered if this might be an example of a phenomenon known as "Minnesota Nice," or keeping up appearances that everything is fine (Jones, 2009). I originally worried that only people who had high scores on SRTs would give permission to access the SRT scores. What bore out was that everyone had high scores on SRTs, so this worry was unfounded. This is known as a positive response bias (Arreola, 2007). It addresses the idea that students will not rate an instructor at the low end of the scale unless something specific happened in the

learning context to upset the students. As discussed above, it makes sense to include interpretive data for comparison purposes. It makes further sense to explore whether some other type of assessment is needed to differentiate levels of teaching effectiveness in order to relate teaching effectiveness to prior training.

Regardless, less than 20% of each score could be explained by the dependent variables. What variables explain the other 80% or more for each question?

Future Research

How do students judge the elements on the SRTs? This bears further research. For example, would it have any bearing on how students judge those elements if they knew that the interpretive data says that a choice of 5 is actually interpreted as "poor"? Is there anything that can be done about the positive response bias?

Also, how are instructors interpreting their scores on SRTs, both at the institution that participated in the investigation and at other institutions? As SRTs are generally part of the review process for tenure and promotion, are the interpretive data being included to help understand what the scores are? Research should be completed to investigate these questions, as it may change how SRT scores are viewed and used. While this issue is not directly related to my research question, it is important in building the link between teaching preparation and SRT scores. For example, if SRTs are being used without the interpretive data, then there is no context for the SRT scores, and they will always look good because of the skewedness of the data. Thus, there is no incentive to learn how to teach, as SRT scores are naturally high. If, however, interpretive data are being included and used with SRT scores, it would become

glaringly obvious that a solution for inadequate teaching, as measured by the SRTs needed to be found.

While the next few questions were not within the scope of this research project, they are likely candidates for immediate address in the near future. Dummy variables were coded in the data set in order to look at whether teaching had an effect on the SRT scores. What would change if the data used included the levels of teaching, or the number of years teaching, instead of the dummy variables? Would anything change if the number of years overall teaching were compared to the number of years at the institution in question, when looking at SRT scores? Further, would taking into account the different levels of training on each of the elements from the original survey have any effect on the analysis?

Another question for future research, especially if any of the above questions are found to be significant, is whether finding training on the elements from the survey (conducting a class, constructing a syllabus, giving feedback, etc.) as part of a graduate degree program has any effect on the regression equations. This would need to be investigated in conjunction with a similar question, regarding finding training on one's own.

Finally, if formal higher education has reached the point it has today, with the majority of people who find themselves teaching never having had any formal training in how to teach, what would happen to higher education if they did have that training? There are institutions that claim teaching is critically important. Those same institutions are coming under attack because of the rising costs of higher education.

Could this research, and other research along these lines, be a source for them to justify their existence by furthering their stated point of difference, excellence in teaching?

References

Ackoff, R. (1996). On learning and the systems that facilitate it. *Reflections 1*(1), 14-24.

Akerlind, G. S. (2003). Growing and developing as a university teacher: Variation in meaning. *Studies in Higher Education 28*(4), 375-390.

Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education, 13*(2), 153-166.

Arreola, R. (2007). *Developing a comprehensive faculty evaluation system: A guide to designing, building, and operating large-scale faculty evaluation systems (3rd ed.).* Bolton, MA: Anker Publishing Company, Inc.

Arthur, W., Tubré, T., Paul, D. S., & Edens, P. S. (2003). Teaching effectiveness: The relationship between reaction and learning evaluation criteria. *Educational Psychology, 23*(3), 275-285. doi: 10.1080/0144341032000060110

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

Austin, A. E., Connelly, M. R., & Colbeck, C. L. (2008). Strategies for preparing integrated faculty: The Center for the Integration of Research, Teaching, and Learning. *New Directions for Teaching and Learning, 113*, 69-81. doi: 10.1002/tl

Austin, A. E., & McDaniels, M. (2006). Using doctoral education to prepare faculty to work within Boyer's four domains of scholarship. *New Directions for Institutional Research, 2006*(129), 51-65. doi: 10.1002/ir.171

Bass, H. (2006). Developing scholars and professionals: The case of mathematics. In Golde, C.M., & Walker, G.E. (Eds.). (2006). *Envisioning the future of doctoral education: Preparing stewards of the discipline* (pp. 101-119). San Francisco, CA: Jossey-Bass.

Beaty, L. (1998). The professional development of teachers in higher education: Structures, methods and responsibilities. *Innovations in Education & Training International*, *35*(2), 99-107.

Becher, T. (1994). The significance of disciplinary differences. *Studies in Higher Education 19*(2), 151-161.

Beran, T., & Violato, C. (2005). Ratings of university teacher instruction: How much do student and course characteristics really matter?. *Assessment and Evaluation in Higher education, 30*(6), 593-601. doi: 10.1080/02602930500260688

Boshier, R. (2009). Why is the scholarship of teaching and learning such a hard sell? *Higher Education Research & Development 28*(1), 1–15. doi: 10.1080/07294360802444321

Bowen, W. G., & Rudenstine, N. L. (1992). *In pursuit of the Ph.D.* Princeton, N.J.: Princeton University Press.

Boyer, E. L. (1990). *Scholarship Reconsidered: Priorities of the Professoriate.* Princeton, N.J.: Princeton University Press.

Brew, A. (2003). Teaching and research: New relationships and their implications for inquiry-based teaching and learning in higher education. *Higher Education Research & Development 22*(1), 3-18.

Brint, S. (2008). The Spellings Commission and the case for professionalizing college teaching. *Academe*, *94*(3), 21-24.

Brown, R.B., & McCartney, S. (1998). The link between research and teaching: Its purpose and implications. *Innovations in Education & Training International*, *35*(1), 117-129.

Brookfield, S. (1990). *The Skillful Teacher: On Technique, Trust, and Responsiveness in the Classroom.* San Francisco, CA: Jossey-Bass.

Bulot, J. J., & Johnson, C. J. (2006). Rewards and costs of faculty involvement in intergenerational service-learning. *Educational Gerontology, 32*, 633-645. DOI: 10.1080/03601270500494121

Campbell, D. T. (1967). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin, 54*, 297-312. Adaptation received as a handout in WHRE 8912 during fall semester, 2008.

Capps, K., Melvin, A., Vallano, D., Weiss, M., Watkins, J., Moslemi, J., et al. (2009). Training tomorrow's environmental problem solvers: An integrative approach to graduate education. *Bioscience*, *59*(6), 514-521. doi:10.1025/bio.2009.59.6.10.

Carnell, E. (2007). Conceptions of effective teaching in higher education: Extending the boundaries. *Teaching in Higher Education 12*(1), 25-40. doi: 10.1080/13562510601102081

Center for Teaching and Learning. (n.d.). *Preparing future faculty*. Retrieved from http://www1.umn.edu/ohr/teachlearn/graduate/pff/

Council of Graduate Schools. (n.d.). *Open doors with a doctorate*. Retrieved from http://www.cgsnet.org/portals/0/pdf/Open_Doors_with_a_Doctorate_BW.pdf

Couper, M. P., & Miller, P. V. (2008). Web survey methods. *Public Opinion Quarterly*, *72*(5), 831-835.

Dall'Alba, G. (2005). Improving teaching: Enhancing ways of being university teachers. *Higher Education Research & Development 24*(4), 361–372. doi: 10.1080/07294360500284771

Deem, R., & Lucas, L. (2006). Learning about research: exploring the learning and teaching/research relationship amongst educational practitioners studying in higher education. *Teaching in Higher Education 11*(1), 1-18. doi: 10.1080/13562510500400040

DeNeef, A. L. (2002). *The preparing future faculty program: What difference does it make?* Retrieved from http://www.aacu.org/pff/PFFpublications/what_difference/general_findings.cfm

Dillman, D. A. (2007). Mail and internet surveys: The tailored design method (2nd ed.). Hoboken, N. J.: John Wiley & Sons, Inc.

Douglas, R. M. (1997). Survival of the fittest? Postgraduate education and the professoriate at the fin de siecle. *Daedalus, 126*(4), 137-151.

Eble, K. E. (1972). *Professors as teachers*. San Francisco, CA: Jossey Bass.

Eble, K. E. (1983). *The aims of college teaching*. San Francisco, CA: Jossey Bass.

Eble, K. E., & McKeachie, W. J. (1985). *Improving undergraduate education through faculty development*. San Francisco, CA: Jossey Bass.

Erickson, S. C. (1984). *The essence of good teaching: Helping students learn and remember what they learn*. San Francisco, CA: Jossey-Bass.

Field, A. (2005). *Discovering statistics using SPSS.* London: Sage Publications.

Fleming, R. A., Bazen, E. F., & Wetzstein, M. E. (2005). Measuring the impact of externalities on College of Agriculture teaching evaluations. *Journal of Agriculture and Applied Economics, 37*(30), 635-645.

Foley, D. (2008). Survey of doctorate recipients. Retrieved from http://www.nsf.gov/statistics/showsrvy.cfm?srvy_CatID=3&srvy_Seri=5

Gaff, J. G., Pruitt-Logan, A. S., Sims, L. B., & Denecke, D. D. (2003). *Preparing future faculty in the humanities and social sciences: A guide for change.* Washington, D.C.: Council of Graduate Schools and the Association of American Colleges and Universities.

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). Boston, MA: Pearson Education, Inc.

Golde, C. M. (2006). Preparing stewards of the discipline. In C. M. Golde & G. E. Walker (Eds.), *Envisioning the future of doctoral education: Preparing stewards of the discipline* (pp. 3-20). San Francisco, CA: Jossey-Bass.

Golde, C. M. (2008). Applying lessons from professional education to the preparation of the professoriate. *New Directions for Teaching and Learning, 113*, 17-25. doi: 10.1002/tl.305

Golde, C., & Dore, T. (2001). At cross purposes: What the experiences of today's doctoral students reveal about doctoral education. Retrieved from ERIC database (ED450628).

Golde, C. M., & Walker, G. E. (Eds.). (2006). Envisioning the future of doctoral education: Preparing stewards of the discipline. San Francisco, CA: Jossey-Bass.

Haggerty, K. D. (2010). Tough love: Professional lessons for graduate students. *American Sociologist, 41*(1), A31-32. doi: 10.1007/s12108-010-9088-8

Haladyna, T. M., & Amrein-Beardsley, A. (2009). Validation of a research-based student survey of instruction in a college of education. *Educational Assessment Evaluation and Accountability, 21*(3), 255-276. doi: 10.1007/s11092-008-9065-8

Hattie, J., & Marsh, H. W. (1996). The relationship between research and teaching: A metaanalysis. *Review of Educational Research, 66*(4), 507-542.

Huber, R. M. (1992). *How Professors Play the Cat Guarding the Cream.* Fairfax, Va.: George Mason University Press.

Jackson, D. L., Teal, C. R., Raines, S. J., Nansel, T. R., Force, R. C., & Burdsal, C. A. (1999). The dimensions of students' perceptions of teaching effectiveness. *Educational and Psychological Measurement, 59,* (4), 580-596. doi: 10.1177/00131649921970035

Jackson, W. K., & Simpson, R. D. (1993). Redefining the role of senior faculty at a research university. *New Directions for Teaching and Learning, 55*(4), 69-79.

Johnson, R. (2000). The authority of the student evaluation questionnaire. *Teaching in Higher Education, 5*(4), 419-434.

Jones, S. (December 14, 2009). The unwritten rules that tell Minnesotans how to be nice. *MPR News*. Retrieved from http://minnesota.publicradio.org/display/web/2009/12/14/jones/

Kahanov, L., Loebsack, A. R., Masucci, M. A., & Roberts, J. (2010). Perspectives on parenthood and working of female athletic trainers in the secondary school and collegiate settings. *Journal of Athletic Training 45*(5), 459–466.

Katz, S. (2006). What has happened to the professoriate?. *Chronicle of Higher Education*, *53*(7), B8-B11.

Kittleson, M. J., & Brown, S. L. (2005). E-mail versus web survey response rates among health education professionals. *American Journal of Health Studies, 20*(1), 7-14.

Knight, P. T. (1998). Professional obsolescence and continuing professional development in higher education. *Innovations in Education & Training International*, *35*(3), 248-256.

Kreber, C. (2006). Setting the context: The climate of university teaching and learning. *New Directions For Higher Education 133*, 5-11. doi: 10.1002/he

LaCelle-Peterson, M. W., & Finkelstein, M. J. (1993). Institutions matter: Campus teaching environments' impact on senior faculty. *New Directions for Teaching and Learning, 55*(4), 21-32.

Lambert, L. M., Tice, S. L., & Featherstone, P. H. (Eds.). (1996). *University teaching: A guide for graduate students*. Syracuse, NY: Syracuse University Press.

Leuddeke, G. (2003). Professionalising teaching practice in higher education: A study of disciplinary variation and 'teaching-scholarship'. *Studies in Higher Education 28*(2), 213-228.

Lindsay, R., Breen, R., & Jenkins, A. (2002). Academic research and teaching quality: The views of undergraduate and postgraduate students. *Studies in Higher Education 27*(3), 309-327. doi: 10.1080/03075070220000699

Lowman, J. (1984). *Mastering the Techniques of Teaching*. San Francisco, CA: Jossey-Bass.

Manfreda, K. L., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. *International Journal of Market Research, 50*(1), 79–104.

Marsh, H. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for further research. *International Journal of Educational Research, 11*, 253-388.

McAlpine, L., Maguire, S., & Lee, M. D. (2005). The pedagogy excellence project: A professor_student team approach to authentic inquiry. *Teaching in Higher Education 10*(3), 355-370. doi: 10.1080/13562510500122255

McIntyre, D. (1997). The profession of educational research. *British Educational Research Journal, 23*(2), 127-140.

McKeachie, W. J., & Svinicki, M. (2006). Teaching tips: Strategies, research, and theory for college and university teachers. Boston: Houghton Mifflin Company.

McLean, M. & Barker, H. (2004). Students making progress and the 'research-teaching nexus' debate. *Teaching in Higher Education 9*(4), 407-419. doi: 10.1080/1356251042000252354

Menachemi, N. (2011) Assessing response bias in a web survey at a university faculty. *Evaluation & Research in Education*, *24*(1), 5-15. DOI: 10.1080/09500790.2010.526205

Moore, S., & Kuol, N. (2005). Students evaluating teachers: Exploring the importance of faculty reaction to feedback on teaching. *Teaching in Higher Education 10*(1), 57-73. doi: 10.1080/1356251052000305534

Mourad, M., Kohlwes, J., Maselli, J., & Auerbach, A. D. (2010). Supervising the supervisors: Procedural training and supervision in internal medicine residency. *Journal of General Internal Medicine, 25*(4), 351-356. DOI: 10.1007/s11606-009-1226-z

National Center for Education Statistics. (n.d.). *About IPEDS*. Retrieved from http://nces.ed.gov/ipeds/about/

National Council for the Accreditation of Teacher Education. (2006). *Professional standards for the accreditation of schools, colleges, and departments of education.* Retrieved May 5, 2007, from http://www.ncate.org/public/standards.asp?ch=4

Olivares, O. J. (2003). A conceptual and analytic critique of student ratings of teachers in the USA with implications for teacher effectiveness and student learning. *Teaching in Higher Education, 8*(2), 233-245. doi: 10.1080/1356251032000052465

O'Meara, K. A., & Rice, R. E. (2005). Faculty priorities reconsidered: Rewarding multiple forms of scholarship. San Francisco, CA: Jossey-Bass.

Parker, J. (2002). A new disciplinarity: Communities of knowledge, learning and practice. *Teaching in Higher Education 7*(4), 373-386.

Pickering, A. M. (2006). Learning about university teaching: Reflections on a research study investigating influences for change. *Teaching in Higher Education 11*(3), 319-335. doi: 10.1080/13562510600680756

Pill, A. (2005). Models of professional development in the education and practice of new teachers in higher education. *Teaching in Higher Education 10*(2), 175-188. doi: 10.1080/1356251042000337936

Preparing Future Faculty. (n.d.a). *The preparing future faculty program* [Brochure]. Washington, D.C.: Preparing Future Faculty National Office.

Preparing Future Faculty. (n.d.b). *Who's involved in PFF*. Retrieved from http://www.preparing-faculty.org/PFFWeb.History.htm#pff1

Ramsden, P. (2003). *Learning to Teach in Higher Education*. New York: RoutledgeFalmer.

Rice, R. E., & Richlin, L. (1993). Scholarly work and professional practice. *New Directions for Teaching and Learning, 54*(3), 71-78.

Richlin, L. (1993a). Graduate education and the U.S. faculty. *New Directions for Teaching and Learning, 54*(3), 3-14.

Richlin, L. (1993b). Openness to a broader view of scholarship. *New Directions for Teaching and Learning, 54*(3), 103-108.

Richlin, L. (1993c). To hear all voices: A broader view of faculty scholarship. *New Directions for Teaching and Learning, 54*(3), 39-46.

Ronkowski, S. A. (1993). Scholarly teaching: Developmental stages of pedagogical scholarship. *New Directions for Teaching and Learning, 54*(3), 79-90.

Ross, D. (1986). Training teaching assistants to use active learning strategies. In *Strategies for Active Teaching and Learning in University Classrooms*, (S.F. Schomberg, Ed.). Minneapolis: University of Minnesota. pp. 47-50.

Rust, C. (2000). Do initial training courses have an impact on university teaching? The evidence from two evaluative studies of one course. *Innovations in Education and Training International 37*(3), 254-262. Retrieved from http://www.tandf.co.uk/journals

Schmid, K. L. (1998). The accreditation of university teachers: an optometric viewpoint. *Clinical and Experimental Optometry 81*(3), pp.104-111.

Schuster, J. H., (1993). Preparing the next generation of faculty: The graduate school's opportunity. *New Directions for Teaching and Learning, 54*(3), 27-38.

Schwill, S. A. (2008). The professional practice of mentoring. *American Journal of Education, 115,* 139-167.

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

Song, K. H. (2006). A conceptual model of assessing teaching performance and intellectual development of teacher candidates: A pilot study in the US. *Teaching in Higher Education 11*(2), 175-190. doi: 10.1080/13562510500527701

Speck, B. W. (2003). The role of doctoral programs in preparing faculty for multiple roles in the academy. *New Directions for Higher Education 2003*(124), 41-55.

Stark-Wroblewski, K. A., Ahrling, R. F., & Brill, F. M. (2007). Toward a more comprehensive approach to evaluating teaching effectiveness: Supplementing student evaluations of teaching with pre–post learning measures. *Assessment & Evaluation in Higher Education, 32*(4), 403–415. doi: 10.1080/02602930600898536

Sweeney, A. D. P., Morrison, M. D., Jarratt, D., & Heffernan, T. (2009). Modeling the constructs contributing to the effectiveness of marketing lecturers. *Journal of Marketing Education, 31*(3), 190-202. doi: 10.1177/0273475309345198

Terpstra, D. E., & Honoree, A. L. (2008). The effects of different teaching, research, and service emphases on individual and organizational outcomes in higher education institutions. *Journal of Education for Business, 84*(3), 169-176.

Trigwell, K. & Shale, S. (2004). Student learning and the scholarship of university teaching. *Studies in Higher Education 29*(4), 523-536.

U.S. Department of Education. (2005). *Secretary Spellings announces new commission on the future of higher education*. Retrieved from http://www.ed.gov/news/pressreleases/2005/09/09192005.html

U.S. Department of Education. (2006). *A test of leadership: Charting the future of higher education.* Retrieved from http://www.ed.gov/about/bdscomm/list/hiedfuture/reports.html

Watkins, K. & Marsick, V. (1993). A framework for the learning organization. In *Sculpting the learning organization* (pp. 3-22). 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, and Associates (Eds.), *Paths to the Professoriate: Strategies for Enriching the Preparation of Future Faculty.* San Francisco, CA: Jossey-Bass.

Appendix A: Survey Instrument

CONSENT STATEMENT
Administration of Faculty Teaching Preparation Survey

You are invited to complete a survey on faculty teaching preparation. You were selected as a possible participant because you were an instructor for at least one class at the institution at which you are currently employed for Fall Semester, 2010. We ask that you read this form and ask any questions you may have before agreeing to complete the survey.

This study is being conducted by Margaret Fitzgerald-Sisk, Department of Organizational Leadership, Policy, and Development, as part of the requirements for the doctor of philosophy degree at the University of Minnesota.

Background Information:
The purpose of this survey is to explore whether there is a relationship between faculty teaching preparation and average scores on student ratings of teaching.

Procedures:
If you agree to participate, you will be asked to complete a short survey. The survey should take about 10 minutes. You will also be asked to allow the principal investigator to request the average scores for the first four questions of the institution's Student Ratings of Teaching for classes you have taught at the institution since 2009.

Confidentiality:
The records of this study will be kept private. In any sort of report we might publish, we will not include any information that will make it possible to identify a subject. Research records will be stored securely, and only researchers will have access to the records.

Voluntary Nature of the Study:
Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the institution at which you are currently employed. If you decide to participate, you are free to withdraw at any time without affecting those relationships.

Contacts and Questions:
The researcher conducting this study is Margaret Fitzgerald-Sisk. You may ask any questions you have now. If you have questions later, you may contact her at [information deleted for privacy]. You may also contact her adviser, Jim Brown, at [information deleted for privacy].

If you have any questions or concerns regarding the study and would like to talk to someone other than the researchers, contact the Research Subjects' Advocate line, [information deleted for privacy].

You may have a copy of this form to keep for your records.

Teaching Preparation Survey

Thank you very much for taking the time to complete this important questionnaire about faculty teaching preparation. This survey will be used to gather information about whether teaching preparation has an effect on SRT scores. The directions for responding are provided with each question.

In order to get comparable data, we will be asking you to refer to the education program you were involved in immediately prior to getting your first job teaching at a college or university, as well as the expectations for your current job teaching at a college or university.

1. While you were participating in the education program you were involved in immediately prior to getting your first job teaching at a college or university, how much training (if any) did you receive for the following?

	None	A small amount	A moderate amount	A large amount
Conducting a class for undergraduates	☐	☐	☐	☐
Constructing a syllabus	☐	☐	☐	☐
Grading student work	☐	☐	☐	☐
Giving feedback to students	☐	☐	☐	☐
Choosing instructional methods for a class (lecture, group work, assignments, etc.)	☐	☐	☐	☐

2. While you were participating in the education program you were involved in immediately prior to getting your first job teaching at a college or university, how did you receive training (if any) for the following?

	I didn't receive training for this	As part of the program I was in	I found it on my own
Conducting a class for undergraduates	☐	☐	☐
Constructing a syllabus	☐	☐	☐
Grading student work	☐	☐	☐
Giving feedback to students	☐	☐	☐
Choosing instructional methods for a class (lecture, group work, assignments, etc.)	☐	☐	☐

3. While you were participating in the education program you were involved in immediately prior to getting your first job teaching at a college or university, how much training did you expect to receive for the following?

	None	A small amount	A moderate amount	A large amount
Conducting a class for undergraduates	☐	☐	☐	☐
Constructing a syllabus	☐	☐	☐	☐
Grading student work	☐	☐	☐	☐
Giving feedback to students	☐	☐	☐	☐
Choosing instructional methods for a class (lecture, group work, assignments, etc.)	☐	☐	☐	☐

4. Now that you teach at a college or university, what amount of training in each of the following do you believe is appropriate for current graduate students?

	None	A small amount	A moderate amount	A large amount
Conducting a class for undergraduates	☐	☐	☐	☐
Constructing a syllabus	☐	☐	☐	☐
Grading student work	☐	☐	☐	☐
Giving feedback to students	☐	☐	☐	☐
Choosing instructional methods for a class (lecture, group work, assignments, etc.)	☐	☐	☐	☐

Comments: _____

5. To what degree do you believe your preparation during graduate school for performing these tasks was sufficient?

	Not at all sufficient	Somewhat sufficient	Very sufficient
Conducting a class for undergraduates	☐	☐	☐
Constructing a syllabus	☐	☐	☐
Grading student work	☐	☐	☐
Giving feedback to students	☐	☐	☐
Choosing instructional methods for a class (lecture, group work, assignments, etc.)	☐	☐	☐

Comments: _____

6. To what degree are the following items important for current graduate students to learn if they plan to become faculty members at colleges or universities in the United States?

	Not at all important	A little important	Moderately important	Very important
Researching in the field	☐	☐	☐	☐
Teaching	☐	☐	☐	☐
Preparing articles for publication	☐	☐	☐	☐
Mentoring students	☐	☐	☐	☐

7. Which of the following are expectations for you in your current job teaching at a college or university? Check all that apply – some items may be subsets of others.

- ☐ Teaching
- ☐ Researching
- ☐ Publishing
- ☐ Service to the college or university community
- ☐ Service to a community other than the college or university at which you are employed
- ☐ Grading papers and/or projects from the classes you teach
- ☐ Meeting with students
- ☐ Meeting with other faculty members
- ☐ External projects
- ☐ Internal projects
- ☐ Constructing syllabi for your classes
- ☐ Designing new classes for students
- ☐ Lecturing
- ☐ Monitoring group work in your classes
- ☐ Mentoring students
- ☐ Handing out SRT (evaluation forms)

8. Do you hold a terminal degree in your field?

- o Yes
- o No

9. Is a terminal degree in your field required for the position you currently hold in teaching at a college or university?

- o Yes
- o No

10. What is your official job title?

- o Professor

- Associate Professor
- Assistant Professor
- Instructor
- Lecturer
- Graduate Assistant
- Teaching Assistant
- Other _____

11. As of the end of 2010, how many years have you taught (at any level)?

- less than one year
- 1 year
- 1.5 to 5 years
- 5.5 to 10 years
- more than 10 years

12. At what levels have you taught? Check all that apply.

- ☐ Preschool
- ☐ K-12
- ☐ Undergraduate
- ☐ Graduate
- ☐ Business/Industry

13. As of the end of 2010, how many semesters have you taught at the institution of higher education at which you are currently employed?

- 1 semester
- 2 semesters (1 year)
- 3-10 semesters (1.5-5 years)
- 11-20 semesters (5.5-10 years)
- more than 20 semesters (more than 10 years)

14. In order to explore whether there is a relationship between faculty teaching preparation and average scores on student ratings of teaching, the researcher asks your specific permission to retrieve aggregate scores for each of the first four questions on the Student Ratings of Teaching for classes you have taught at this institution since 2009. With the understanding that confidentiality will be preserved, do you grant your permission?

- Yes
  - Please provide your name: _____
  - Please provide your institutional e-mail address: _____
- No

Appendix B: General Results Tables

Table 18

Means and Standard Deviations for Main Effect Variables by Question

	Q1 N=176		Q2 N=176		Q3 N=176		Q4 N=176	
	M	SD	M	SD	M	SD	M	SD
Questions	5.5488	.39026	5.2540	.57529	5.2056	.56195	5.6374	.29364
1	5.4896	.44396	5.1636	.65171	5.0711	.63654	5.5719	.30095
2	5.4867	.42359	5.1397	.61784	5.0998	.61231	5.5832	.31753
3	5.5117	.45987	5.1577	.71988	5.0881	.71094	5.5822	.33956
4	5.5300	.45748	5.1771	.66754	5.1053	.67363	5.6144	.31615
5	5.4583	.45149	5.1202	.63127	5.0625	.62719	5.5825	.31678
6	5.0954a	.37008a	4.5592a	.57955a	4.9677a	.48888a	5.5638a	.24690a
	5.3663b	.41979b	5.0013b	.59152b	4.9788b	.42963b	5.5475b	.23921b
	5.4959c	.36129c	5.1964c	.51910c	5.1387c	.58035c	5.5762c	.29796c
	5.6829d	.21448d	5.5014d	.30347d	5.4541d	.42331d	5.7443d	.23302d
	5.5913e	.43620e	5.2715e	.63774e	5.1361e	.61587e	5.6199e	.32833e
7	5.5671	.27311	5.5929	.26056	5.3900	.36014	5.6643	.28442
8	5.6589	.31723	5.4441	.42924	5.4159	.39093	5.7123	.27055
9	5.5405	.38897	5.2554	.55848	5.2096	.54294	5.6396	.28205
10	5.6014	.38271	5.2894	.56306	5.1540	.57800	5.6398	.32228
11	5.6162	.36648	5.3667	.43255	5.2010	.52620	5.6817	.23149
12	5.1507a	.37812a	4.6627a	.60442a	5.0107a	.50095a	5.5620a	.23337a
	5.4579b	.35801b	5.1867b	.48105b	5.1100b	.46097b	5.5950b	.22409b
	5.5774c	.30347c	5.3274c	.45626c	5.2866c	.51252c	5.6464c	.30716c
	5.5968d	.40237d	5.3421d	.54941d	5.3550d	.53092d	5.7179d	.23449d
	5.6498e	.42130e	5.3276e	.65277e	5.1339e	.66668e	5.6251e	.34603e

Note. Row headings: 1 – TraininRcvd_ConductClass; 2 – TrainingRcvd_ConstrSyll; 3 – TrainingRcvd_Grading; 4 – TrainingRcvd_GiveFeedback; 5 – TrainingRcvd_ChooseInstrMeth; 6 – YearsTaughtAnyLevel; 7 – TaughtPreschool; 8 – TaughtK-12; 9 – TaughtUndergrad; 10 – TaughtGrad; 11 – TaughtBusIndust; 12 – YearsTaughtThisInstitution.
a=less than 1 year; b=1 year; c=1.5 to 5 years; d=5.5 to 10 years; e=more than 10 years

Table 19

Correlations of Variable of Interest by Question

1	2	3	4	5	6	7	8	9	10	11	12
Q1: Instructor is well-prepared (mean) N = 176											
1.000											
-.128											
-.187	.530*										
-.072	.634*	.405*									
-.041	.625*	.460*	.778*								
.235	.671*	.493*	.535*	.648*							
.312*	.195	-.023	.147	.214	.054						
.010	.006	-.062	-.154	-.171	-.031	.124					
.163	-.060	-.253*	-.164	-.220	-.269*	.195	.084				
-.078	-.093	-.048	-.124	-.093	-.042	-.054	-.060	-.104			
.129	.142	.146	.105	.211	.103	.456*	.097	-.105	-.057		
.097	.178	.018	.076	.124	.154	.306*	-.114	.015	.046	.266*	
.297*	.150	.009	.134	.259	.113	.828*	.022	-.052	.039	.500*	.324*
Q2: Instructor presents clearly N = 176											
1.000											
-.133											
-.234	.530*										
-.127	.634*	.405*									
-.113	.625*	.460*	.778*								
-.236	.671	.493*	.535*	.648*							
.287*	.195	-.023	.147	.214	.054						
.120	.006	-.062	-.154	-.171	-.031	.124					
.191	-.060	-.253*	-.164	-.220	-.269*	.195	.084				
.009	-.093	-.048	-.124	-.093	-.042	-.054	-.060	-.104			
.059	.142	.146	.105	.211	.103	.456*	.097	-.105	-.057		
.110	.178	.018	.076	.124	.154	.306*	-.114	.015	.046	.266*	
.229	.150	.009	.134	.259	.113	.828*	.022	-.052	.039	.500*	.324*

Q3: Instructor Provides Feedback (mean) N = 176

1.000											
-.202											
-.222	.530*										
-.159	.634*	.405*									
-.151	.625*	.460*	.778*								
-.258	.671*	.493*	.535*	.648*							
.092	.195	-.023	.147	.214	.054						
.067	.006	-.062	-.154	-.171	-.031	.124					
.217	-.060	-.253*	-.164	-.220	-.269	.195	.084				
.026	-.093	-.048	-.124	-.093	-.042	-.054	-.060	.104			
-.088	.142	.146	.105	.211	.103	.456*	.097	-.105	-.057		
-.005	.178	.018	.076	.124	.154	.306*	-.114	.015	.046	.266*	
.040	.150	.009	.134	.259	.113	.828*	.022	-.052	.039	.500*	.324*

Q4: Instructor Presented Clearly (mean) N = 176

1.000											
-.133											
-.234	.530*										
-.127	.634*	.405*									
-.113	.625*	.460*	.778*								
-.236	.671*	.493*	.535*	.648*							
.287*	.195	-.023	.147	.214	.054						
.120	.006	-.062	-.154	-.171	-.031	.124					
.191	-.060	-.253	-.164*	-.220	-.269*	.195	.084				
.009	-.093	-.048	-.124	-.093	-.042	-.054	-.060	-.104			
.059	.142	.146	.105	.211	.103	.456*	.097	-.105	-.057		
.110	.178	.018	.076	.124	.154	.306*	-.114	.015	.046	.266*	
.229	.150	.009	.134	.259*	.113	.828*	.022	-.052	.039	.500*	.324*

Note. * $p < .001$. Column and row headings: 1 – question under consideration (Q1, Q2, Q3, Q4); 2 – TraininRcvd_ConductClass-None; 3 – TrainingRcvd_ConstrSyll; 4 – TrainingRcvd_Grading; 5 – TrainingRcvd_GiveFeedback; 6 – TiningRcvd_ChooseInstrMeth-None; 7 – YearsTaughtAnyLevel; 8 – TaughtPreschool; 9 – TaughtK-12; 10 – TaughtUndergrad; 11 – TaughtGrad; 12 – TaughtBusIndust; 13 – YearsTaughtThisInstitution.

Appendix C: SRT Question One Results Tables

Table 20

Casewise Diagnostics for Question One

Case Number	Std. Residual	InstrWellPrep-Mean	Predicted Value	Residual
3	-2.372	4.86	5.7129	-.85293
5	-2.981	4.53	5.6016	-1.07158
7	-3.565	4.32	5.6016	-1.28158
19	-2.451	4.62	5.5012	-.88118
35	-2.700	4.43	5.4008	-.97078
39	-2.118	4.84	5.6016	-.76158
67	-2.225	4.40	5.2000	-.79998
78	-3.787	4.24	5.6016	-1.36158
89	-2.785	4.50	5.5012	-1.00118
116	-2.204	4.82	5.6125	-.79253
167	-2.063	4.67	5.4117	-.74172

Note. a. Dependent Variable: InstrWellPrep-Mean

Table 21

Case Summaries for Question One

Case Number	Cook's Distance	Mahalanobis Distance	Centered Leverage Value	Standardized DFBETA Intercept	Standardized DFBETA TrainingInstrMeth
5	.05693	2.24648	.01284	.23306	-.20045
7	.08143	2.24648	.01284	.28210	-.24264
19	.02438	1.08608	.00621	.05645	-.17922
35	.03093	1.17750	.00673	-.08384	-.21709
39	.02876	2.24648	.01284	.16342	-.14056
67	.06187	5.11584	.02923	-.31530	-.21327
78	.09192	2.24648	.01284	.30128	-.25914
89	.03148	1.08608	.00621	.06448	-.20472
116	.01956	1.06886	.00611	-.14988	.16658
167	.04557	4.29195	.02453	-.36834	.12967
Total	11				

Casewise Diagnostics

I examined the summary table of residual statistics for extreme cases. Any cases that have a standardized value of less than -2 or greater than 2 appear in Table 20 in this appendix. In an ordinary sample, we would expect 95% of the cases to have standardized residuals within about ±2. The sample has 176 data points, so it is reasonable to expect about nine (5%) cases to have standardized residuals outside of these limits. For question one, this is exactly what we have. In addition, 99% of cases should lie within ±2.5, and so we would expect only 1% of cases to lie outside these limits. From the cases listed in Table 20, it is clear that five cases (2.8%) lie outside of the limits (cases 5, 7, 35, 78, and 89). Therefore, our sample appears to conform to what we would expect for a fairly accurate model. There is no real cause for concern,

except for cases 7 and 78, which have standardized residuals greater than 3, which was large enough for me to investigate further.

Table 21 (Case Summaries) shows the influence statistics for the cases in Table 20. None of the cases have a Cook's distance greater than 1, and the two cases of concern, 7 and 78, have Cook's distances well below this value. Therefore, none of the cases is having an undue influence on the model. The average leverage is calculated as .017 (k+1/n = 3/176), so we are looking for Centered Leverage Values that are twice as large (.034). All the cases are within this boundary. Finally, the guidelines for the Mahalanobis distance indicate that with a sample of 100, values greater than 15 are cause for concern. None of the Mahalanobis distances approach this value. Thus we can conclude that there are no influential cases within our data.

Finally, the DFBeta statistics show whether a case would have a large influence on the regression parameters by presenting an absolute value greater than one. All of the DFBeta statistics lie within ±1 for question two, so none of these cases have an undue influence over the regression parameters.

Table 22

Multiple Regression Coefficients for Q1

	B	*SE B*	*β*
Step 1			
Constant	5.40	0.15	
TrainConduct Class	-0.04	0.09	-0.04
TrainingConstrSyll	-0.06	0.07	-0.07
TrainingGrading	-0.02	0.10	-0.02
TrainingFeedback	0.11	0.10	0.14
TrainInstrMethods	-0.21	0.09	-0.26*
YrsTaughtAll	0.03	0.05	-0.10
TaughtPreschool	-0.01	0.15	-0.01
TaughtK-12	0.07	0.08	0.08
TaughtUndergrad	-0.13	0.11	-0.09
TaughtGrad	-0.01	0.07	-0.02
TaughtBusIndust	0.03	0.07	0.03
YrsTaughtU	0.07	0.04	0.22
Step 2			
Constant	5.404	0.15	
TrainConduct Class	-0.04	0.09	-.05
TrainingConstrSyll	-0.05	0.07	-.07
TrainingGrading	-0.02	0.10	-.03
TrainingFeedback	0.11	0.10	.14
TrainInstrMethods	-0.21	0.08	-.26*
YrsTaughtAll	0.03	0.05	.10
TaughtK-12	0.07	0.08	.08
TaughtUndergrad	-0.13	0.11	-.09
TaughtGrad	-0.01	0.07	-.02
TaughtBusIndust	0.03	0.07	.04
YrsTaughtU	0.07	0.04	.22
Step 3			
Constant	5.41	0.15	
TrainConduct Class	-0.04	0.09	-.05
TrainingConstrSyll	-0.06	0.07	-.07
TrainingGrading	-0.02	0.10	-.03
TrainingFeedback	0.11	0.10	.14
TrainInstrMethods	-0.21	0.08	-.26*
YrsTaughtAll	0.03	0.05	.10
TaughtK-12	0.07	0.07	.08
TaughtUndergrad	-0.13	0.11	-.08
TaughtBusIndust	0.03	0.07	.03
YrsTaughtU	0.07	0.04	.22

Step 4			
Constant	5.40	0.15	
TrainConduct Class	-0.04	0.09	-.05
TrainingConstrSyll	-0.06	0.07	-.07
TrainingFeedback	0.10	0.08	.12
TrainInstrMethods	-0.20	0.08	-.26*
YrsTaughtAll	0.03	0.05	.09
TaughtK-12	0.07	0.07	.08
TaughtUndergrad	-0.13	0.11	-.08
TaughtBusIndust	0.03	0.07	.03
YrsTaughtU	0.07	0.04	.22
Step 5			
Constant	5.39	0.15	
TrainConduct Class	-0.04	0.09	-.05
TrainingConstrSyll	-0.06	0.07	-.07
TrainingFeedback	0.10	0.08	.12
TrainInstrMethods	-0.20	0.08	-.26*
YrsTaughtAll	0.03	0.05	.10
TaughtK-12	0.08	0.07	.08
TaughtUndergrad	-0.13	0.11	-.08
YrsTaughtU	0.07	0.04	.23
Step 6			
Constant	5.40	0.15	
TrainingConstrSyll	-0.07	0.07	-.09
TrainingFeedback	0.09	0.08	.11
TrainInstrMethods	-0.22	0.08	-.28*
YrsTaughtAll	0.03	0.05	.09
TaughtK-12	0.07	0.07	.08
TaughtUndergrad	-0.12	0.11	-.08
YrsTaughtU	0.07	0.04	.24
Step 7			
Constant	5.44	0.14	
TrainingConstrSyll	-0.07	0.07	-.08
TrainingFeedback	0.08	0.08	.11
TrainInstrMethods	-0.22	0.08	-.28**
TaughtK-12	0.09	0.07	.10
TaughtUndergrad	-0.13	0.11	-.09
YrsTaughtU	0.10	0.02	.31***
Step 8			
Constant	5.40	0.13	
TrainingFeedback	0.07	0.08	.09
TrainInstrMethods	-0.24	0.07	-.30**
TaughtK-12	0.10	0.07	.11
TaughtUndergrad	-0.13	0.11	-.08
YrsTaughtU	0.10	0.02	.32***

Step 9			
Constant	5.40	0.13	
TrainInstrMethods	-0.19	0.06	-.25**
TaughtK-12	0.09	0.07	.10
TaughtUndergrad	-0.14	0.11	-.09
YrsTaughtU	0.10	0.02	.33***
Step 10			
Constant	5.27	0.08	
TrainInstrMethods	-0.19	0.06	-.24**
TaughtK-12	0.10	0.07	.12
YrsTaughtU	0.10	0.02	.33***
Step 11			
Constant	5.31	0.08	
TrainInstrMethods	-0.21	0.06	-.27***
YrsTaughtU	0.10	0.02	.33***

Note. $R^2 = .20$ for Step 1, $\Delta R^2 = .00$ for Step 2, $\Delta R^2 = .00$ for Step 3, $\Delta R^2 = .00$ for Step 4, $\Delta R^2 = -.001$ for Step 5, $\Delta R^2 = .00$ for Step 6, $\Delta R^2 = -.002$ for Step 7, $\Delta R^2 = -.005$ for Step 8, $\Delta R^2 = -.004$ for Step 9, $\Delta R^2 = -.008$ for Step 10, $\Delta R^2 = -.012$ for Step 11.
*$p < .05$. **$p < .01$. ***$p < .001$.

Appendix D: SRT Question Two Results Tables

Table 23

Casewise Diagnostics for Question Two

Case Number	Std. Residual	InstrPresClear-Mean	Predicted Value	Residual
3	-2.773	4.14	5.6154	-1.47543
5	-4.446	2.88	5.2455	-2.36547
7	-3.149	3.57	5.2455	-1.67547
19	-2.773	3.77	5.2455	-1.47547
67	-2.142	3.70	4.8397	-1.13967
78	-3.770	3.24	5.2455	-2.00547
116	-3.280	3.43	5.1750	-1.74500
167	-2.299	3.67	4.8929	-1.22291

Note. a. Dependent Variable: InstrPresClear-Mean

Table 24

Case Summaries for Question Two

Case Number	Cook's Distance	Mahalanobis Distance	Centered Leverage Value	Standardized DFBETA Intercept	Standardized DFBETA TrainingConstrSyll	Standardized DFBETA TrainingInstrMeth	Standardized DFBETA YrsTaught	Standardized DFFIT
3	.04082	2.57045	.01469	.03202	.15670	.12392	-.21337	-.41240
5	.08826	2.02379	.01156	.33631	-.17051	-.21048	-.33996	-.63046
7	.04428	2.02379	.01156	.23072	-.11697	-.14439	-.23322	-.43251
19	.03434	2.02379	.01156	.20177	-.10230	-.12628	-.20396	-.37824
67	.09135	11.08216	.06333	-.47242	.36000	-.34881	.44806	-.61154
78	.06344	2.02379	.01156	.27996	-.14194	-.17521	-.28300	-.52482
116	.07678	3.73498	.02134	-.19247	-.39754	.41866	.14893	-.57125
167	.07959	8.44253	.04824	-.41511	-.26410	.27343	.41753	-.57195
Total	8							

Casewise Diagnostics

We must examine the summary table of residual statistics for extreme cases. Any cases that have a standardized value of less than -2 or greater than 2 appear in the table below. In an ordinary sample, we would expect 95% of the cases to have standardized residuals within about ±2. We have a sample of 176, so it is reasonable to expect about nine (5%) cases to have standardized residuals outside of these limits. For question two, there are fewer than nine cases that fall outside these boundaries. In addition, 99% of cases should lie within ±2.5, and so we would expect only 1% of cases to lie outside these limits. From the cases listed here, it is clear that six cases (3.4%) lie outside of the limits (cases 3, 5, 7, 19, 78, and 116). Therefore, our sample appears to conform to what we would expect for a fairly accurate model. There is no real cause for concern, except for cases 5, 7, 78, and 116, which have standardized residuals greater than 3, which is large enough for us to investigate further.

Table 24 (Case Summaries) shows the influence statistics for the cases in Table 21. None of the cases have a Cook's distance greater than 1, and the six cases of concern have Cook's distances well below this value. Therefore, none of the cases is having an undue influence on the model. The average leverage is calculated as .017 (k+1/n = 3/176), so we are looking for Centered Leverage Values that are twice as large (.034). Most of the potential outliers have values within this boundary. Only cases 67 and 167 do not. Still, these cases have standardized residuals, as mentioned above, that fall within the ±2.5 boundary, so there is likely not a cause for concern. Finally, the guidelines for the Mahalanobis distance indicate that with a sample of 100, values greater than 15 are cause for concern. None of the Mahalanobis distances are greater

than 15. Thus we can conclude that there are no influential cases within our data for question two.

Finally, the DFBeta statistics show whether a case would have a large influence on the regression parameters by presenting an absolute value greater than one. All of the DFBeta statistics lie within ±1 for question two, so none of these cases have an undue influence over the regression parameters.

Table 25

Multiple Regression Coefficients for Q2

	B	*SE B*	*β*
Step 1			
Constant	4.89	0.23	
TrainConduct Class	-0.01	0.14	-.01
TrainingConstrSyll	-0.13	0.11	-.11
TrainingGrading	-0.05	0.15	-.04
TrainingFeedback	0.10	0.16	.09
TrainInstrMethods	-0.24	0.13	-.21
YrsTaughtAll	0.09	0.07	.20
TaughtPreschool	0.29	0.23	.10
TaughtK-12	0.09	0.11	.07
TaughtUndergrad	0.02	0.17	.01
TaughtGrad	-0.08	0.10	-.07
TaughtBusIndust	0.11	0.11	.08
YrsTaughtU	0.04	0.07	.09
Step 2			
Constant	4.90	0.22	
TrainingConstrSyll	-0.13	0.10	-.11
TrainingGrading	-0.05	0.14	-.05
TrainingFeedback	0.10	0.16	.09
TrainInstrMethods	-0.25	0.12	-.22*
YrsTaughtAll	0.09	0.07	.20
TaughtPreschool	0.29	0.22	.10
TaughtK-12	0.09	0.11	.07
TaughtUndergrad	0.02	0.17	.01
TaughtGrad	-0.08	0.10	-.07
TaughtBusIndust	0.11	0.11	.08
YrsTaughtU	0.04	0.07	.09
Step 3			
Constant	4.92	0.16	
TrainingConstrSyll	-0.13	0.10	-.11
TrainingGrading	-0.06	0.14	-.05
TrainingFeedback	0.10	0.15	.09
TrainInstrMethods	-0.25	0.12	-.22*
YrsTaughtAll	0.09	0.07	.19
TaughtPreschool	0.29	0.22	.10
TaughtK-12	0.09	0.11	.07
TaughtGrad	-0.08	0.10	-.07
TaughtBusIndust	0.11	0.10	.08
YrsTaughtU	0.04	0.06	.09
Step 4			
Constant	4.92	0.15	

TrainingConstrSyll	-0.13	0.10	-.12
TrainingFeedback	0.06	0.12	.05
TrainInstrMethods	-0.25	0.12	-.22*
YrsTaughtAll	0.09	0.07	.19
TaughtPreschool	0.30	0.22	.10
TaughtK-12	0.09	0.11	.07
TaughtGrad	-0.08	0.10	-.07
TaughtBusIndust	0.11	0.10	.08
YrsTaughtU	0.04	0.06	.10
Step 5			
Constant	4.90	0.15	
TrainingConstrSyll	-0.12	0.10	-.11
TrainInstrMethods	-0.22	0.10	-.19*
YrsTaughtAll	0.09	0.07	.20
TaughtPreschool	0.27	0.21	.09
TaughtK-12	0.09	0.11	.07
TaughtGrad	-0.08	0.10	-.07
TaughtBusIndust	0.10	0.10	.08
YrsTaughtU	0.05	0.06	.10
Step 6			
Constant	4.91	0.15	
TrainingConstrSyll	-0.13	0.10	-.11
TrainInstrMethods	-0.22	0.10	-.19*
YrsTaughtAll	0.13	0.04	.27**
TaughtPreschool	0.25	0.21	.08
TaughtK-12	0.06	0.10	.05
TaughtGrad	-0.06	0.10	-.06
TaughtBusIndust	0.11	0.10	.08
Step 7			
Constant	4.91	0.15	
TrainingConstrSyll	-0.14	0.10	-.12
TrainInstrMethods	-0.23	0.09	-.20*
YrsTaughtAll	0.14	0.04	.29**
TaughtPreschool	0.25	0.21	.09
TaughtGrad	-0.08	0.09	-.07
TaughtBusIndust	0.11	0.10	.08
Step 8			
Constant	4.94	0.15	
TrainingConstrSyll	-0.15	0.09	-.13
TrainInstrMethods	-0.22	0.09	-.20*
YrsTaughtAll	0.13	0.04	.26**
TaughtPreschool	0.24	0.21	.08
TaughtBusIndust	0.10	0.10	.07
Step 9			
Constant	4.92	0.15	
TrainingConstrSyll	-0.15	0.09	-.13

TrainInstrMethods	-0.21	0.09	-.18*
YrsTaughtAll	0.14	0.03	.29***
TaughtPreschool	0.21	0.21	.07
Step 10			
Constant	4.91	0.15	
TrainingConstrSyll	-0.16	0.09	-.13*
TrainInstrMethods	-0.21	0.09	-.18*
YrsTaughtAll	0.14	0.03	.29***

Note. $R^2 = .18$ for Step 1, $\Delta R^2 = .00$ for Step 2, $\Delta R^2 = .00$ for Step 3, $\Delta R^2 = .00$ for Step 4, $\Delta R^2 = -.001$ for Step 5, $\Delta R^2 = -.002$ for Step 6, $\Delta R^2 = -.002$ for Step 7, $\Delta R^2 = -.003$ for Step 8, $\Delta R^2 = -.004$ for Step 9, $\Delta R^2 = -.005$ for Step 10.
*$p < .05$. **$p < .01$. ***$p < .001$.

Appendix E: SRT Question Three Results Tables

Table 26

Casewise Diagnostics for Question Three

Case Number	Std. Residual	InstrProvFeedback-Mean	Predicted Value	Residual
5	-3.552	3.12	5.0348	-1.91479
7	-2.365	3.76	5.0348	-1.27479
35	-2.012	3.95	5.0348	-1.08479
48	-2.423	3.97	5.2766	-1.30656
61	-2.545	4.11	5.4818	-1.37185
78	-3.515	3.14	5.0348	-1.89479
89	-2.235	3.83	5.0348	-1.20479
116	-2.794	3.77	5.2766	-1.50656
159	-2.272	3.81	5.0348	-1.22479

Note. a. Dependent Variable: InstrProvFeedback-Mean

Table 27

Case Summary Data for Question Three

Case Number	Cook's Distance	Mahalanobis Distance	Centered Leverage Value	Standardized DFBETA Intercept	Standardized DFBETA TrainingInstrMeth	Standardized DFBETA TaughtK-12
5	.05094	1.07590	.00615	-.04782	-.24288	.09075
7	.02258	1.07590	.00615	-.03115	-.15823	.05912
35	.01635	1.07590	.00615	-.02639	-.13402	.05007
48	.03217	1.79058	.01023	-.31525	.22819	.16613
61	.05569	3.30190	.01887	-.08197	.11324	-.30307
78	.04988	1.07590	.00615	-.04728	-.24014	.08972
89	.02017	1.07590	.00615	-.02939	-.14927	.05577
116	.04277	1.79058	.01023	-.36567	.26468	.19270
159	.02084	1.07590	.00615	-.02989	-.15182	.05673
Total	9					

Casewise Diagnostics

We must examine the summary table of residual statistics for extreme cases. Any cases that have a standardized value of less than -2 or greater than 2 appear in the table below. In an ordinary sample, we would expect 95% of the cases to have standardized residuals within about ±2. We have a sample of 176, so it is reasonable to expect about nine (5%) cases to have standardized residuals outside of these limits. For question three, there are exactly nine cases that fall outside these boundaries. In addition, 99% of cases should lie within ±2.5, and so we would expect only 1% of cases to lie outside these limits. From the cases listed here, it is clear that four cases (2%) lie outside of the limits (cases 5, 61, 78, and 116). Therefore, our sample appears to conform to what we would expect for a fairly accurate model. There is no real cause for concern, except for cases 5 and 78, which have standardized residuals greater than 3, which is large enough for us to investigate further.

Table 27 (Case Summaries) shows the influence statistics for the cases in Table 24. None of the cases have a Cook's distance greater than 1, and the four cases of concern have Cook's distances well below this value. Therefore, none of the cases is having an undue influence on the model. The average leverage is calculated as .017 (k+1/n = 3/176), so we are looking for Centered Leverage Values that are twice as large (.034). All of the potential outliers have values within this boundary. Finally, the guidelines for the Mahalanobis distance indicate that with a sample of 100, values greater than 15 are cause for concern. None of the Mahalanobis distances are greater

than 15. Thus we can conclude that there are no influential cases within our data for question three.

Finally, the DFBeta statistics show whether a case would have a large influence on the regression parameters by presenting an absolute value greater than one. All of the DFBeta statistics lie within ±1 for question three, so none of these cases have an undue influence over the regression parameters.

Table 28

Multiple Regression Coefficients for Q3

	B	*SE B*	*β*
Step 1			
Constant	5.10	0.23	
TrainConduct Class	-0.11	0.14	-.10
TrainingConstrSyll	-0.07	0.11	-.06
TrainingGrading	-0.06	0.15	-.05
TrainingFeedback	0.16	0.16	.14
TrainInstrMethods	-0.21	0.13	-.18
YrsTaughtAll	0.04	0.07	.08
TaughtPreschool	0.20	0.23	.07
TaughtK-12	0.18	0.11	.14
TaughtUndergrad	0.06	0.17	.03
TaughtGrad	-0.14	0.10	-.12
TaughtBusIndust	0.04	0.11	.03
YrsTaughtU	0.02	0.07	.04
Step 2			
Constant	5.10	0.23	
TrainConduct Class	-0.12	0.14	-.10
TrainingConstrSyll	-0.07	0.11	-.06
TrainingGrading	-0.06	0.15	-.05
TrainingFeedback	0.16	0.16	.14
TrainInstrMethods	-0.21	0.13	-.18
YrsTaughtAll	0.05	0.04	.11
TaughtPreschool	0.19	0.23	.07
TaughtK-12	0.17	0.11	.13
TaughtUndergrad	0.06	0.17	.03
TaughtGrad	-0.13	0.10	-.12
TaughtBusIndust	0.04	0.11	.03
Step 3			
Constant	5.16	0.16	
TrainConduct Class	-0.12	0.14	-.10
TrainingConstrSyll	-0.07	0.11	-.07
TrainingGrading	-0.06	0.14	-.06
TrainingFeedback	0.16	0.16	.14
TrainInstrMethods	-0.20	0.13	-.18
YrsTaughtAll	0.05	0.04	.11
TaughtPreschool	0.19	0.22	.06
TaughtK-12	0.16	0.11	.13
TaughtGrad	-0.14	0.10	-.12
TaughtBusIndust	0.04	0.11	.03
Step 4			
Constant	5.16	0.16	

TrainConduct Class	-0.11	0.14	-.10
TrainingConstrSyll	-0.08	0.10	-.07
TrainingGrading	-0.07	0.14	-.06
TrainingFeedback	0.16	0.15	.14
TrainInstrMethods	-0.20	0.13	-.18
YrsTaughtAll	0.06	0.04	.12
TaughtPreschool	0.17	0.22	.06
TaughtK-12	0.16	0.11	.13
TaughtGrad	-0.13	0.10	-.12
Step 5			
Constant	5.15	0.16	
TrainConduct Class	-0.13	0.13	-.11
TrainingConstrSyll	-0.08	0.08	-.07
TrainingFeedback	0.11	0.11	.10
TrainInstrMethods	-0.19	0.19	-.17
YrsTaughtAll	0.06	0.06	.12
TaughtPreschool	0.17	0.17	.06
TaughtK-12	0.17	0.17	.13
TaughtGrad	-0.13	0.13	-.11
Step 6			
Constant	5.11	0.15	
TrainConduct Class	-0.16	0.12	-.14
TrainingFeedback	0.10	0.12	.09
TrainInstrMethods	-0.20	0.12	-.18
YrsTaughtAll	0.06	0.04	.13
TaughtPreschool	0.18	0.22	.06
TaughtK-12	0.18	0.10	.14
TaughtGrad	-0.13	0.09	-.12
Step 7			
Constant	5.10	0.15	
TrainConduct Class	-0.15	0.12	-.13
TrainingFeedback	0.08	0.12	.07
TrainInstrMethods	-0.19	0.12	-.17
YrsTaughtAll	0.07	0.04	.14
TaughtK-12	0.18	0.10	.14
TaughtGrad	-0.13	0.09	-.11
Step 8			
Constant	5.09	0.15	
TrainConduct Class	-0.12	0.12	-.11
TrainInstrMethods	-0.17	0.12	-.15
YrsTaughtAll	0.07	0.04	.15
TaughtK-12	0.17	0.10	.13
TaughtGrad	-0.12	0.09	-.11
Step 9			
Constant	5.11	0.14	
TrainInstrMethods	-0.25	0.09	-.22**

YrsTaughtAll	0.06	0.04	.13
TaughtK-12	0.15	0.10	.12
TaughtGrad	-0.13	0.09	-.11
Step 10			
Constant	5.15	0.14	
TrainInstrMethods	-0.25	0.09	-.23**
YrsTaughtAll	0.04	0.04	.08
TaughtK-12	0.18	0.10	.14
Step 11			
Constant	5.28	0.07	
TrainInstrMethods	-0.24	0.08	-.22**
TaughtK-12	0.21	0.10	.16*

Note. $R^2 = .12$ for Step 1, $\Delta R^2 = .00$ for Step 2, $\Delta R^2 = .00$ for Step 3, $\Delta R^2 = .00$ for Step 4, $\Delta R^2 = -.001$ for Step 5, $\Delta R^2 = -.003$ for Step 6, $\Delta R^2 = -.003$ for Step 7, $\Delta R^2 = -.002$ for Step 8, $\Delta R^2 = -.006$ for Step 9, $\Delta R^2 = -.010$ for Step 10, $\Delta R^2 = -.006$ for Step 11.
*$p < .05$. **$p < .01$. ***$p < .001$.

Appendix F: SRT Question Four Results Tables

Table 29

Casewise Diagnostics for Question Four

Case Number	Std. Residual	InstrRespect-Mean	Predicted Value	Residual
5	-3.001	4.71	5.5638	-.85380
7	-2.649	4.81	5.5638	-.75380
35	-2.751	4.90	5.6826	-.78264
47	-2.551	5.00	5.7258	-.72582
60	-2.398	4.83	5.5123	-.68227
61	-2.375	5.05	5.7258	-.67582
78	-3.493	4.57	5.5638	-.99380
111	-2.094	5.13	5.7258	-.59582
116	-3.374	4.54	5.5001	-.96009

Note. a. Dependent Variable: InstrRespect-Mean

Table 30

Case Summaries for Question Four

Case Number	Cook's Distance	Mahalanobis Distance	Centered Leverage Value	Standardized DFBETA Intercept	Standardized DFBETA TrainingConstrSyll	Standardized DFBETA TrainingGrading	Standardized DFBETA TrainingInstrMeth	Standardized DFBETA TrainingGiveFeedback
5	.03099	1.91732	.01096	.06107	-.05443	-.14151	-.04044	-.02162
7	.02415	1.91732	.01096	.05358	-.04776	-.12416	-.03548	-.01897
35	.09649	8.93401	.05105	.02540	-.10265	.61140	-.06987	-.51059
47	.02010	1.62803	.00930	-.32238	.14581	.01415	.07319	.00301
60	.04663	5.57817	.03188	.01513	-.16635	.06658	-.32327	.25558
61	.01743	1.62803	.00930	-.29938	.13541	.01314	.06797	.00279
78	.04198	1.91732	.01096	.07181	-.06400	-.16639	-.04755	-.02543
111	.01355	1.62803	.00930	-.26293	.11892	.01154	.05969	.00245
116	.21965	13.24932	.07571	-.11403	-.40922	-.88571	.34062	.70871
Total	9							

Casewise Diagnostics

We must examine the summary table of residual statistics for extreme cases. Any cases that have a standardized value of less than -2 or greater than 2 appear in the table below. In an ordinary sample, we would expect 95% of the cases to have standardized residuals within about ±2. We have a sample of 176, so it is reasonable to expect about nine (5%) cases to have standardized residuals outside of these limits. For question four, there are exactly nine cases that fall outside these boundaries. In addition, 99% of cases should lie within ±2.5, and so we would expect only 1% of cases to lie outside these limits. From the cases listed here, it is clear that six cases (3.4%) lie outside of the limits (cases 5, 7, 35, 47, 78, and 116). Therefore, our sample appears to conform to what we would expect for a fairly accurate model. There is no real cause for concern, except for cases 5, 78, and 116, which have standardized residuals greater than 3, which is large enough for us to investigate further.

Table 30 (Case Summaries) shows the influence statistics for the cases in Table 29. None of the cases have a Cook's distance greater than 1, and the six cases of concern have Cook's distances well below this value. Therefore, none of the cases is having an undue influence on the model. The average leverage is calculated as .017 (k+1/n = 3/176), so we are looking for Centered Leverage Values that are twice as large (.034). Most of the potential outliers have values within this boundary. Case 35 has a Centered Leverage Value of .05105, which is outside the boundary. Yet, all the other summary data for this case indicate that it is likely not exerting undue influence. Finally, the guidelines for the Mahalanobis distance indicate that with a sample of 100, values greater than 15 are cause for concern. None of the Mahalanobis distances are

greater than 15. Thus we can conclude that there are no influential cases within our data for question four.

Finally, the DFBeta statistics show whether a case would have a large influence on the regression parameters by presenting an absolute value greater than one. All of the DFBeta statistics lie within ±1 for question four, so none of these cases have an undue influence over the regression parameters.

Table 31

Multiple Regression Coefficients for Q4

	B	*SE B*	*β*
Step 1			
Constant	5.63	0.12	
TrainConduct Class	-0.08	0.07	-.14
TrainingConstrSyll	-0.07	0.06	-.12
TrainingGrading	-0.09	0.08	-.05
TrainingFeedback	0.18	0.08	.29
TrainInstrMethods	-0.08	0.07	-.13
YrsTaughtAll	0.01	0.04	.04
TaughtPreschool	0.06	0.12	.04
TaughtK-12	0.07	0.06	.10
TaughtUndergrad	0.03	0.09	.02
TaughtGrad	-0.01	0.05	-.02
TaughtBusIndust	0.07	0.06	.10
YrsTaughtU	-0.002	0.04	-.01
Step 2			
Constant	5.63	0.12	
TrainConduct Class	-0.08	0.07	-.14
TrainingConstrSyll	-0.07	0.06	-.12
TrainingGrading	-0.09	0.08	-.15
TrainingFeedback	0.17	0.08	.29
TrainInstrMethods	-0.08	0.07	-.14
YrsTaughtAll	0.01	0.02	.04
TaughtPreschool	0.06	0.12	.04
TaughtK-12	0.07	0.06	.10
TaughtUndergrad	0.02	0.09	.02
TaughtGrad	-0.01	0.05	-.02
TaughtBusIndust	0.07	0.06	.10
Step 3			
Constant	5.63	0.12	
TrainConduct Class	-0.08	0.07	-.14
TrainingConstrSyll	-0.07	0.06	-.13
TrainingGrading	-0.09	0.08	-.15
TrainingFeedback	0.17	0.08	.29*
TrainInstrMethods	-0.08	0.07	-.13
YrsTaughtAll	0.01	0.02	.03
TaughtPreschool	0.05	0.12	.04
TaughtK-12	0.07	0.06	.11
TaughtUndergrad	0.03	0.09	.02
TaughtBusIndust	0.07	0.06	.10
Step 4			

Constant	5.66	0.08	
TrainConduct Class	-0.08	0.07	-.14
TrainingConstrSyll	-0.08	0.06	-.13
TrainingGrading	-0.09	0.08	-.15
TrainingFeedback	0.17	0.08	.29*
TrainInstrMethods	-0.08	0.07	-.13
YrsTaughtAll	0.01	0.02	.03
TaughtPreschool	0.05	0.12	.03
TaughtK-12	0.07	0.06	.10
TaughtBusIndust	0.07	0.06	.10
Step 5			
Constant	5.68	0.04	
TrainConduct Class	-0.08	0.07	-.13
TrainingConstrSyll	-0.08	0.05	-.13
TrainingGrading	-0.09	0.08	-.15
TrainingFeedback	0.18	0.08	.30*
TrainInstrMethods	-0.08	0.07	-.14
TaughtPreschool	0.06	0.11	.04
TaughtK-12	0.07	0.05	.11
TaughtBusIndust	0.08	0.05	.11
Step 6			
Constant	5.68	0.04	
TrainConduct Class	-0.07	0.07	-.12
TrainingConstrSyll	-0.08	0.05	-.13
TrainingGrading	-0.09	0.08	-.16
TrainingFeedback	0.17	0.08	.29*
TrainInstrMethods	-0.07	0.07	-.13
TaughtK-12	0.07	0.05	.11
TaughtBusIndust	0.07	0.05	.10
Step 7			
Constant	5.69	0.04	
TrainingConstrSyll	-0.09	0.05	-.16
TrainingGrading	-0.12	0.07	-.19
TrainingFeedback	0.17	0.08	.28*
TrainInstrMethods	-0.11	0.06	-.18
TaughtK-12	0.06	0.05	.09
TaughtBusIndust	0.06	0.05	.09
Step 8			
Constant	5.71	0.04	
TrainingConstrSyll	-0.10	0.05	-.17*
TrainingGrading	-0.12	0.07	-.19
TrainingFeedback	0.16	0.08	.28*
TrainInstrMethods	-0.12	0.06	-.20
TaughtBusIndust	0.07	0.05	.10
Step 9			

Constant	5.73	0.04	
TrainingConstrSyll	-0.11	0.05	-.18*
TrainingGradin	-0.12	0.07	-.20
TrainingFeedback	0.17	0.08	.29*
TrainInstrMethods	-0.11	0.06	-.18

Note. $R^2 = .11$ for Step 1, $\Delta R^2 = .00$ for Step 2, $\Delta R^2 = .00$ for Step 3, $\Delta R^2 = .00$ for Step 4, $\Delta R^2 = .00$ for Step 5, $\Delta R^2 = -.001$ for Step 6, $\Delta R^2 = -.006$ for Step 7, $\Delta R^2 = -.007$ for Step 8, $\Delta R^2 = -.009$ for Step 9. $^*p < .05$, $^{**}p < .01$, and $^{***}p < .001$.

Appendix G: Survey and SRT Variables Mapped

Table 32

Variable Map

Questions	Variable
Q1: The instructor was well prepared for class.	InstrWellPrep-Mean
Q2: The instructor presented the subject matter clearly.	InstrPresClear-Mean
Q3: The instructor provided feedback intended to improve my course performance.	InstrProvFeedback-Mean
Q4: The instructor treated me with respect.	InstrRespect-Mean
While you were participating in the education program you were involved in immediately prior to getting your first job teaching at a college or university, how much training (if any) did you receive for the following?	TraininRcvd_ConductClass-None TrainingRcvd_ConstrSyll TrainingRcvd_Grading TrainingRcvd_GiveFeedback TiningRcvd_ChooseInstrMeth-None
As of the end of 2010, how many years have you taught (at any level)?	YearsTaughtAnyLevel
At what levels have you taught?	TaughtPreschool TaughtK-12 TaughtUndergrad TaughtGrad TaughtBusIndust
As of the end of 2010, how many semesters have you taught at the institution under consideration?	YearsTaughtThisInstitution