STUDENT RATINGS OF INSTRUCTION AND STUDENT MOTIVATION:
IS THERE A CONNECTION?

by

CHRISTOPHER R. FEIT

B.S., University of Wisconsin-Eau Claire, 2002
M.S., Oklahoma State University-Stillwater, 2005

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Special Education, Counseling and Student Affairs
College of Education

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2014

Abstract

This study examined factors relates to student ratings of instruction and student levels of motivation. Data came from archival data of 386,195 classes of faculty and students who completed the *Faculty Information Form* (FIF), completed by the instructor, and the *Student Ratings Diagnostic Form* (SRDF) completed by the student from the Individual Development and Educational Assessment (IDEA) Center Student Ratings system. Descriptive statistics, correlation studies, analysis of variance (ANOVA), and pairwise comparisons were used to test the research hypotheses. Despite significant differences among student ratings of instruction and student motivation by course type, discipline, and student type, the amount of unknown variability in student ratings of instruction and student motivation is still very large. The findings from the study provide higher education institutions with information about differences between student ratings of instruction by institution type, course level, discipline, and course type as well as the impact of student motivation on student ratings of instruction.

STUDENT RATINGS OF INSTRUCTION AND STUDENT MOTIVATION: IS THERE A CONNECTION?

by

CHRISTOPHER R. FEIT

B.S., University of Wisconsin-Eau Claire, 2002
M.S., Oklahoma State University-Stillwater, 2005

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Special Education, Counseling and Student Affairs
College of Education

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2014

Approved by:

Major Professor
Doris W. Carroll

Abstract

This study examined factors relates to student ratings of instruction and student levels of motivation. Data came from archival data of 386,195 classes of faculty and students who completed the *Faculty Information Form* (FIF), completed by the instructor, and the *Student Ratings Diagnostic Form* (SRDF) completed by the student from the Individual Development and Educational Assessment (IDEA) Center Student Ratings system. Descriptive statistics, correlation studies, analysis of variance (ANOVA), and pairwise comparisons were used to test the research hypotheses. Despite significant differences among student ratings of instruction and student motivation by course type, discipline, and student type, the amount of unknown variability in student ratings of instruction and student motivation is still very large. The findings from the study provide higher education institutions with information about differences between student ratings of instruction by institution type, course level, discipline, and course type as well as the impact of student motivation on student ratings of instruction.

Table of Contents

List of Figures ix

List of Tables x

Acknowledgements xi

Dedication xii

Chapter 1 - INTRODUCTION TO THE TOPIC 1

Purpose of the Study 4

Brief Student Ratings of Instruction Literature Review 5

Statement of the Problem 7

Significance of the Study 7

Summary 8

Chapter 2 - REVIEW OF THE LITERATURE 9

Introduction 9

Reliability and Validity 10

Reliability of Student Ratings of Instruction 10

Validity of Student Ratings of Instruction 13

Faculty Concerns Over using Student Ratings of Instruction 18

Comparisons by Academic Rank 20

Course Level 23

Summary Related to Research of Student Ratings of Instruction 24

Student Motivation 25

Humanistic 26

Behaviorism 26

Cognitive 27

Social Cognitive Theory 28

Reciprocal Determinism 29

Self-Observation 30

Self-Evaluation 30

Self-Reaction 31

Self-Efficacy 31
Performance Outcomes 32
Vicarious Experiences 32
Verbal Persuasion 33
Physiological Feedback 33
The Effects of Self-Efficacy 34
Magnitude 34
Generality 35
Strength 35
Limitations Regarding Self-Efficacy 35
Self-Efficacy and Student Learning 36
Research Questions 38
Chapter 3 - METHODS 39
Introduction 39
Quasi-Experimental Design 39
Validity concerns of the Experimental Design 40
Participants 41
Instrumentation 42
Reliability of IDEA Center Instrumentation 44
Validity of IDEA Center Instrumentation 45
Statistical Procedures 50
Limitation of the Study 51
Summary 53
Chapter 4 - RESULTS 54
Introduction 54
Descriptive Statistics for Independent Variables 54
Degree Granting Type of Institution 54
Course Type 55
Discipline 56
Student Type 58
Cross Tabulation by Degree Granting Type of Institution 59

Research Question 1 59
Research Question 2 63
Research Question 3 67
Research Question 4 71
Research Question 5 75
Summary 76
Chapter 5 - DISCUSSSION 77
Purpose of the Study 77
Summary of the Results 77
Discussion of Research Questions 78
Research Question 1 Discussion: 78
Research Question 2 Discussion: 80
Research Question 3 Discussion: 82
Research Question 4 Discussion: 84
Research Question 5 Discussion: 86
Considerations for Future Research 86
Overall Recommendations for Practice 87
Summary 90
References 91
Appendix A - IDEA Center Forms 104
Appendix B - List of STEM CIP Codes 111
Appendix C - STATISTICAL TABLES 126

List of Figures

Figure 1 *Triadic Reciprocal Determinism* 29
Figure 2 *Non-Doctoral: Lower Level General Education Teaching Effectiveness* 62
Figure 3 *Non-Doctoral: Lower Level General Education Teaching Effectiveness* 62
Figure 4 *Doctoral: Lower Level General Education Teaching Effectiveness* 66
Figure 5 *Doctoral: Upper Level General Education Teaching Effectiveness* 66
Figure 6 *Non-Doctoral: Lower Level General Education Student Motivation* 70
Figure 7 *Non-Doctoral: Upper Level General Education Student Motivation* 70
Figure 8 *Doctoral: Lower Level General Education Student Motivation* 74
Figure 9 *Doctoral: Upper Level General Education Student Motivation* 74

List of Tables

Table 3.1 *SRDF Item Reliabilities: Split Half Reliabilities and Standard Errors*........................ 45
Table 3.2 *Inter-Correlations of IDEA Faculty Information Form (FR) and IDEA Diagnostic Form (SR)*... 47
Table 3.3 *Relationship of Teaching Methods to Learning Objectives*.. 48
Table 3.4 *Differences Between Adjusted and Unadjusted Ratings Among Five Types of Classes*49
Table 4.1 *Highest Degree Offered: IDEA Center Classifications*.. 55
Table 4.2 *Primary Teaching Method: Classifications from Faculty Information Form (FIF)* 55
Table 4.3 *Discipline Counts as Defined by Researcher* 58
Table 4.4 *Principal Type of Student Enrolling in This Course* 58
Table 4.5 *Cross Tabulation by Degree Granting Type of Institution*.. 59
Table 4.6 *Three-Way Analysis of Variance Output for Teaching Effectiveness at non-doctoral Institutions*... 60
Table 4.7 *Three-Way Analysis of Variance Output for Teaching Effectiveness at Doctoral Institutions*... 64
Table 4.8 *Three-Way Analysis of Variance Output for Student Motivation at non-doctoral Institutions*... 68
Table 4.9 *Three-Way Analysis of Variance Output for Student Motivation at Doctoral Institutions*... 72

Acknowledgements

I would like to thank the members of my doctoral committee for the encouragement, assistance, and patience during my degree program. Without the help and guidance offered by Dr. Doris W. Carroll, Dr. Christy Craft, Dr. Sheryl Hodge, and Dr. Leigh Murray I would not have been able to complete my doctoral program. I would like to personally thank Dr. Doris W. Carroll for taking me on as an advisee after my original advisor left Kansas State University.

I would also like to acknowledge the support and encouragement offered by my wife and children. My wife Christine for putting up with me during weekend writing binges, my children for understanding that "Daddy is writing" and all of them for keeping me motivated to finish.

I would also like to thank several people that believed in me throughout my academic career and pushed me to always strive for more, specifically: Jodi Thesing-Ritter, Dr. Duane Oyen, Dr. Matthew Waters, Dr. Marcia Dickman (late), Dr. Dale Fuqua, Dr. Bob Huss, Dr. Kelline (Kelli) Cox and most importantly my mother. My mom always valued the importance of education and took great personal sacrifice so I could pursue higher education. I truly am grateful for the opportunity provided by mother, thank you Nancy Feit.

Lastly I would like to thank Dr. Stephen Benton and the members of the Individual Development and Educational Assessment (IDEA) Center staff in Manhattan, KS. The IDEA Center is a non-profit organization dedicated to the advancement of learning in higher education. Primarily through the use of student ratings of instruction and administrator evaluations, they provide assistance to institutions across the United States. Without their help and access to their database, I would not have been able to produce this study.

Dedication

I dedicate this dissertation to my paternal grandparents, the late Robert A. Feit Sr. and Maryanne Feit, and my maternal grandparents, Robert (late) and Antoinette Liebelt. My grandparents have always been an important part of my life. I truly cherished the time I spent with them and the memories we shared. Thanks for always making me feel special. Looking forward to seeing all of you again soon.

Chapter 1 - INTRODUCTION TO THE TOPIC

Student Ratings of Instruction (SRI) and Student Evaluations of Teaching (SET) are two terms or concepts prevalent in the literature of higher education and educational psychology. A keyword search for articles in in the Educational Research Information Clearinghouse (ERIC) database for the phrase “student evaluations of teaching” yielded 42,324 hits while a similar search for student ratings of instruction yielded 3,346 hits. When adding higher education to the search function, the results are reduced to 22,300 and 1,968 respectively. Clearly, the interest in SRI/SET is relevant and well documented.

So what makes SRI so important? Student ratings of instruction are used as the primary tool to measure teaching effectiveness in higher education (Huemer, 2005). Marsh (1984; 1987) listed the four primary uses of SRI as: (1) feedback on teaching to faculty, (2) data to be used in personnel decisions, (3) assistance for students in course selection, and (4) data for research purposes. It is clear that SRI, and the research and discussion related to SRI, are an important topic within higher education for the foreseeable future. The question becomes how to effectively use SRI data in higher education planning and student learning.

Using SRI along with motivational strategies encompasses a holistic approach to student learning, which is the cornerstone of higher education in modern society. Student ratings of instruction and student motivational theory may seem disconnected, but when considering Locke’s (1968) Theory of Goal Attainment and Bandura’s (1977; 1989; 2001) Social Cognitive Theory, specifically self-efficacy, the connection is more noticeable. Both SRI and Goal Attainment and self-efficacy, are focused on outcomes, and in relation to this study, student outcomes. Student ratings of instruction seek to determine if a meaningful learning experience occurred. Bandura's self-efficacy construct, as well as Locke's Theory of Goal Attainment incorporate these meaningful learning experiences to meet the societal, employer, or personal goals or outcomes for the individual.

Locke's (1968) Theory of Goal Commitment is best summarized by task performance is dependent on the goals that individuals set for themselves on a given task (Locke & Latham, 1990). The crux of the theory involves the willingness of the individual to set SMART goals. SMART goals are: (1) specific, (2) measurable, (3) attainable, (4) realistic, and (5) time-bound (Locke, 1968; Locke & Latham, 2002)). If goals are not framed around these SMART principles, the goal-setting will not be effective and therefore limit, perhaps prohibit, the likelihood of success.

Originally proposed by Bandura in 1977, self-efficacy, the most significant part of Bandura's Social Cognitive Theory, is the belief that an individual has about his/her ability to do well (Bandura, 1977; 1986; Graham & Weiner, 1996). A central component of self-efficacy is the idea of goal attainment. Locke and Latham (1990; 2002) argued that self-efficacy, specifically positive self-efficacy, is the primary factor in achieving one's goal. According to Bandura (1977) self-efficacy is the central question one must answer prior to undertaking any task. Since we do not usually attempt tasks where we expect to fail (Lunenburg, 2011), having a strong self-efficacy belief structure is needed for individuals to seek higher learning at a college or university setting.

The core basis for self-efficacy judgments come from (1) performance outcomes or past experiences (most critical), (2) vicarious experiences, or how one measures up to self-determined peers, (3) verbal persuasion/encouragement, and (4) interpretations of physiological feedback (nerves) (Bandura, 2001). Once an individual makes a self-efficacy judgment, the self-efficacy effect is dictated by three factors. The first factor, the magnitude, is the level of difficulty required to complete the task. The generality factor is the transferability of the task to a broader sense. Lastly, the strength of the effect is the willingness of the individuals to complete tasks in the face of obstacles (Bandura, 1977; Lunenburg, 2011).

Bandura's self-efficacy construct is widely documented in a variety of applications. Self-efficacy has been used in counseling, coaching, health-care, and more relevant to this motivation study. Because of the transferability of self-efficacy techniques across such a diverse set of disciplines, using self-efficacy for student motivation the rationale for using it is credible and strong (Bandura, 1977; Graham & Weiner, 1996).

Unlike the two motivational models briefly referenced above, student ratings of instruction were developed most likely out of desperation (Costin et al, 1971). Faculty and researchers alike have stated, and it seems like an obvious conclusion, that the best way to measure student learning is to ask students. Many institutions used in-house questionnaires that were institution, college, or even department specific in the beginning (Costin et al, 1971). Although the questionnaires were easy to develop and administer, the psychometric properties were modest at best (Marsh, 1984). And since each institution, college, or department had a separate form or list of questions, comparisons across departments/colleges were difficult.

Standardized SRI forms were developed by the 1980s. The Individual Development and Educational Assessment (IDEA) form from Kansas State University (Cashin & Slawson, 1977) and the Student Evaluation of Educational Quality (SEEQ) form developed by Marsh (1982, 1984) were two of the highly referenced standardized SRI, although several other rating forms were used by institutions across the United States. With the advent of standardized SRI forms, the concerns over the psychometric properties of the in-house questionnaires were minimized

(Aleamoni, 1999; d'Apollonia & Abrami, 1997; Feldman, 1978, 1984, 1987, 2007; Greenwald, 1997; Marsh, 1987, 1991, 2001; Wright & Jenkins-Guarnieri, 2012). Despite the relative strength of using standardized SRI, issues and questions over the use of SRI remained, and based on the amount of recent literature on the subject, the importance of the topic has not waned.

Purpose of the Study

The purpose of this study was to examine institution type, student type, discipline, and course type and to determine how these variables, both individually and collectively, affect student ratings of instruction. While student ratings of instruction and differences among student ratings of instruction have been examined, the issue of institution type (public vs. private) and degrees awarded (bachelor vs. doctoral) remain relatively unexplored (Centra, 2009). One of the first requirements for a prospective student, with the help of a professional or faculty advisor, is to choose a series of classes based on their academic or professional career goals. Especially early in their academic career, exposure to different academic disciplines may have long-term educational ramifications, regardless if the experience was positive or negative. A better understanding of SRI differences among institution type may help explain why students' rate similar courses differently among institution type or instructor type (Educational Testing Service, 2010a; 2010b).

Understanding institution type, course type, discipline, and student type can improve department and institution comparisons. As institutions move towards a regulated and data-informed governance structure, having accurate and timely data to guide decisions is an imperative. In addition, with the heightened level of outside public scrutiny on higher education, the pressure to be able to document and defend differences between institutions is critical. Higher education faculty and administrators must first understand the differences before they can defend the differences. In order to understand the differences higher education administrators need to have a better understanding of what preceded this study and how theory can add weight and credibility to this study.

Brief Student Ratings of Instruction Literature Review

The history of SRI/SET research began in the 1960s and 1970s. The literature focused on the reliability and validity of student ratings of instruction. If institutions were using SRI at all, they were in-house evaluations primarily that were often unreliable and had difficulty adequately measuring or capturing the characteristics of effective teaching (Marsh, 1984).

One of the most referenced studies is the infamous Dr. Fox Lecture (Peer & Babad, 2012), which was a classic study in educational seduction by Naftulin, Ware, and Donnelly (1973).

The Dr. Fox study was an experimental design in an effort to disprove the reliability of student ratings of instruction. The term experimental design is used loosely because of the numerous criticisms of the methodology (Marsh, 1987, 2007; Theall & Franklin, 2001). The criticisms aside, the Dr. Fox Lecture disguised a lecture loaded with inaccurate information and double talk by using a vibrant lecturer and entertaining discussion. Despite the ruse, Naftulin, Ware, and Donnelly (1973) found a statistically significant likelihood that the lecture provided a positive and meaningful instructional activity. In a more recent study, Peer and Babad (2012) conducted a replication of the Dr. Fox study and confirmed that students enjoyed an entertaining lecture and were impressed by instructors with high credentials but when asked directly about learning, no seduction of learning occurred.

Following the Dr. Fox fallout, several researchers attempted to demonstrate the reliability, validity, and practicality of student ratings of instruction (Centra, 1978; Feldman, 1978; Marsh, 1984, 1987). The issues of reliability of SRI were focused around consistency, stability, and generalizability. Feldman (1984) and Marsh (1987) concluded that as long as you have an adequate number of raters (more than 10), the consistency of student ratings were solid. Similarly, Marsh (2007) demonstrated SEEQ results over a 13 year period were stable. Lastly, the generalizability of SRI, assuming a valid number of observations from several sections, including multiple teaching modes (lecture, discussion, lab, etc.), have been relatively undisputed (Gillmore et al, 1978; Marsh, 1984; Marsh & Overall, 1981; Murray et al, 1990).

Unlike the reliability of SRI, issues of validity in SRI have remained today. Ory and Ryan (2001) summarized the research relating to SRI validity around five issues: (1) multisection, (2) multitrait/multimethod, (3) bias/prejudice, (4) experimenter variance, and (5) conceptual structure. Although multisection (d'Apolloinia & Abrami, 1997; Greenwald, 1997; Greenwald & Gillmore, 1997), multitrait/multimethod (Marsh, 1982), and conceptual structure (Feldman, 1976; Marsh, 1987; 1991) issues are relatively settled, concerns over bias such as instructor/student gender, class type, and grading leniency, are still debated in the literature (d'Apolloinia & Abrami, 1997; Feldman, 1976; Greenwald, 1997; Greenwald & Gillmore, 1997; Marsh, 1982; 1987; 1991). Experimenter variance studies are primarily interested in the Dr. Fox Lecture (Naftulin, Ware, & Donnelly, 1973) study.

Two of the first nationally normed and psychometrically tested instruments were Cashin and Slawson's (1977) Individual Development and Educational Assessment (IDEA) rating form and Marsh's (1984) Students' Evaluation of Educational Quality (SEEQ). With the advent of these standardized SRI, the literature regarding SRI shifted from reliability and validity and more towards issues of bias (Aleamoni, 1999; Centra, 2009; Feldman, 2007; Oliver-Hoyo, 2008).

Concerns regarding SRI and issues of institution type, course level, and faculty demographics are the focus of the current literature related to student ratings of instruction.

Statement of the Problem

According to Seldin (1993) SRI have become the most used technique to measure teaching effectiveness in higher education. Despite the popularity of SRI with students and university administration, decisions based on SRI data are often made by individuals or groups who have an insufficient understanding of SRI data. Specifically, the desire of committees and administrators to compare ratings of instruction across disciplines and course types (McKeachie, 1997).

Additionally, students are becoming attentive to not only course selection, but to instructor and institution type. Students find themselves debating about the cost difference between a traditional four-year institution and a community college/technical school. A student's decisions to attend an institution, or even enroll in a course, are often a result of many factors. Which characteristics of instructor and institution type are most conducive to specific academic aims/goals? How do instructor type and course level affect student learning? Why do student ratings of instruction (SRI) differ by instructor type? These and other related questions are the focus of this study.

Significance of the Study

The results of this study could have implications for providing academic administrators, department heads, curriculum coordinators, faculty, and even current and potential students with information about what characteristics affect student learning regarding course level, discipline, and institution size. Specifically, this study focuses on institution type, course level, and discipline, and student type. The findings from this study provide insight into the discrepancies between student ratings among differences in course level, instructor and student type.

Summary

Student ratings of instruction have a definite and growing purpose in higher education. Despite the efforts of faculty and even some administrators to discredit or dispute the use of some SRI to measure effectiveness of teaching and learning (Culver, 2010; El Hassan, 2009; Huemer, 2005), their use will continue to grow with continued calls for teaching accountability by governing boards and colleges and universities needing to find, document, and honor good teachers (Feldman, 2007). Therefore, the need for further examination regarding differences in SRI and how motivation may or may not affect teaching effectiveness is warranted. The next chapter presents a full review of self-efficacy, the student development theory relevant to this study, and the psychometric properties of student ratings of instruction.

Chapter 2 - REVIEW OF THE LITERATURE

Introduction

The literature of student ratings of instruction (SRI) and student evaluations of teaching is abundant. Summarizing all the literature regarding SRI is beyond the scope of any thesis or dissertation. However, for the purposes of this study, and reasons of relevance, the following literature review summarized the research related to the psychometric properties of SRI, research related to differences among SRI by course type, institution type, and the theoretical framework student motivational theory. Like stated earlier, the literature specific to these parameters is abundant. Therefore, only seminal works with more recent supporting studies were included in this literature review.

Because of the importance of student learning in any institute of higher education, evaluating teaching is an important component in improving the student learning experience. The most popular and common method of evaluating teaching is using student ratings of instruction. The use of SRI is only likely to increase in the future due to the increased emphasis institutions are putting on good teaching (Feldman, 2007). According to Huemer (2005), SRIs are used for three main reasons: (1) reduced cost of administration (as compared to other methods), (2) perceived objectivity, and (3) limited alternatives. In addition, student ratings of instruction have been used for: (1) feedback on teaching, (2) administrative decision making, (3) aid in student decision making on course selection, and (4) data for research on teaching (Marsh & Dunkin, 1992). Because of the popularity of student ratings of instruction and the decisions that are based on them such as teaching assignments, tenure and promotion to name a few, there has been considerable scrutiny over the reliability and validity of student ratings of instruction.

Reliability and Validity

Reliability and validity are two commonly cited, but often confused, constructs in the social sciences. The simple definition of reliability is the ability for a technique, procedure, or observation to reveal like results each time it is performed. Conversely, validity is the ability of the collection instrument to measure the proper construct, feeling, or belief (Agresti & Finlay, 2009). Neither concept is interchangeable, nor does demonstrating reliability prove validity.

Reliability of Student Ratings of Instruction

The reliability of student ratings of instruction has been demonstrated by numerous researchers and cited throughout the literature (Greenwald & Gillmore, 1997; Huemer, 2005; Marsh 1984; Marsh & Roche, 1997; 2000). Although positive associations between learning and student ratings of instruction existed, the relationship varied across disciplines and appeared strongest in education and liberal art courses (Clayson, 2008). However, the literature debating the reliability of student ratings of instruction does exist.

The classic piece of literature, and often cited as the reason for studies on the reliability of student ratings of instruction, is the infamous “Dr. Fox Lecture” (Naftulin, Ware & Donnelly, 1973). The Dr. Fox Lecture is one of the few research studies where opponents of using SRI/SET had empirical data from an experimental study disputing the usefulness of SRI (Peer & Babad, 2012). The general hypothesis for Naftulin, Ware & Donnelly’s (1973) work was if students were adequately distracted in an alternative learning situation or style, they would feel that they had a significant learning experience even if the total situation was a ruse. In the study, Dr. Fox was trained by researchers in the subject matter, given a counterfeit curriculum vita, and presented as a guest lecturer. The teaching experience was meant to include contradictory statements. The lecture was administered to three test groups, (1.) a group including psychiatrists, psychologists, and social work educators, (2.) mental health educators, and (3.) graduate students in an educational philosophy course. The post-lecture questionnaire rated the experience as significantly positive. In one instance, respondents stated that they had read some of Dr. Fox's publications. Naftulin, Ware, & Donnelly (1973) stated they were surprised that no one identified the true purpose of the experience considering the academic background and training of the research subjects.

The reliability of an educational measurement tool generally focuses on three main sub-categories: consistency, stability (over time), and generalizability. Thousands of studies on the reliability of student ratings of instruction have been conducted, but in general, they can all be categorized into these three sub-categories.

Consistency in educational measurement is defined as the uniformity of successive results or events (Benton & Cashin, 2012). In other words, how accurate are the observations or ratings within a given collection period? Consistency related to student ratings was most commonly studied as the agreement between classmates on a single item or series of items. Researchers referred to this agreement as interrater agreement (Benton & Cashin, 2012; Streiner, 2003). The interrater agreement was measured by a reliability coefficient (identified usually with an r = some value) with values from .00 (weak) to 1.00 (strong). As related to student ratings of instruction, there was a positive relationship between consistency and class size or as class size increases the overall rating of instructional quality increases. Marsh (1987) found that reliability of Students Evaluation of Educational Quality (SEEQ) factors to range from a low of r = .23 from a class of one evaluation to a high of r = .95 for 50 plus evaluations. Although the range is large, even a class size of 10 evaluations generated a reliability coefficient of .74 on the SEEQ factors. In other words, as the number of raters for a class or observation group increases, the reliability coefficients increase. Zhao and Gallant (2012) found the reliability coefficient for the Ohio State questionnaire to be .95; only about five percent of the variability in student responses are due to error. The same philosophy held for multiple class ratings compared to single class observations (Feldman, 1984; Marsh, 1987).

Stability with regards to student ratings of instruction has dealt with consistency of the ratings over time. Most studies of this type focused on a single instructor, or a group of instructors, ratings over time. Using SEEQ data from more than 50,000 classes collected over a 13 year period Marsh (2007) concluded that teaching effectiveness based on student ratings of instruction were highly stable over time and stable across course levels. Very few instructors showed any significant change in effectiveness over time. Overall, the literature confirmed that student ratings of instruction were stable over time (Bausell et al, 1975; Centra 1993; Marsh 1984; Murray et al, 1990).

Arguably, generalizability is the most important component of reliability. Generalizability addresses an instructor's overall teaching ability in a variety of courses and class types as opposed to a specific situation or course (Benton & Cashin, 2012). For student ratings of instruction to be effective and meaningful, they need to measure teaching effectiveness for a variety of teaching styles and academic subjects independent of a specific course. Since Marsh and Overall (1981) concluded that student ratings of instruction reflected the instructor's effectiveness as opposed to the particular course effect(s), obtaining multiple sets of ratings would be prudent in making any decisions regarding teaching effectiveness (Gillmore et al, 1978; Murray et al, 1990). Marsh (1984) stated that due to the high likelihood of an instructor teaching multiple courses over his/her teaching career, scores from as many classes should be used to make better conclusions about the instructor's effectiveness. No studies dispute Marsh's findings related to the generalizability of student ratings of instruction.

Despite obvious outliers, the reliability of student ratings of instruction are mostly universally accepted (Wright & Jenkins-Guranieri, 2012). In fact, very few researchers spend time debating the reliability of student ratings of instruction, especially the Individual Development and Educational Assessment (IDEA) and SEEQ forms that are used across institutions of higher education. Murray et al (1990) summed up the research on reliability of student ratings of instruction nicely by stating, "Although findings are sometimes contradictory, the weight of evidence suggests that student ratings of a given instructor are reasonably stable across items, raters, and time periods (p. 250)." So, if we agree that student ratings are reliable, we must also explore if they are valid.

Validity of Student Ratings of Instruction

If reliability is defined as the ability to reproduce the same, or similar, results, validity is the degree to which the measurement accurately reflects the intended purpose (Agresti & Finlay, 2009). With student ratings of instruction, the purpose is to capture the effectiveness of teaching. Scholars have long debated what constitutes effective teaching. Marsh, Feldman, and others have conducted numerous studies on effective teaching, and although they have come up with similar results, no one unified set of traits or factors have been identified for effective teaching. Instead, the validity studies of student ratings of instruction are framed around five themes. Multisection, Multitrait/Multimethod, Bias/Prejudice, Experimenter Variance, and Conceptual structure are the five most common measures of validity in student ratings of instruction (Ory & Ryan, 2001).

Multisection validity deals with the relationship between student ratings of instruction and student achievement over several sections of the same course taught by different instructors

(Ory & Ryan, 2001). Richardson (2005) concluded that there is little to no relationship between the ratings given by students taking the same course taught by different teachers and that ratings are based on teacher performance and not a specific subject or class. Although moderately valid over multiple sections of the same course, course characteristics, type of instructor, and type of course may have influences on student ratings of instruction (d'Apolloinia & Abrami, 1997). However, effective teaching consistently has produced a high correlation between student ratings and student learning via student grades (Greenwald, 1997; Greenwald & Gillmore, 1997). In addition, ratings were more valid when they evaluated full-time faculty in large class sections (d'Apolloinia & Abrami, 1997).

Multitrait-Multimethod validity involves the pairing of student ratings of instruction with outside measures of effective instruction (Ory & Ryan, 2001). Specific examples would include alumni ratings, peer ratings, or faculty peer ratings. These studies focused on the effect of convergent and discriminant validity. Marsh (1982) defined the two as, convergent validity or the correlation between factors by two different groups and discriminant validity as the uniqueness of each evaluation factors. Marsh (1982), using the Campbell Fiske (1959) technique, paired student ratings of instruction with faculty self-evaluations to demonstrate convergent validity and the partial discriminant validity of student ratings of instruction.

Bias/Prejudice studies on validity focused on non-teaching related influences on student ratings. A commonly cited definition of bias provided by Centra (2009) stated, "Bias exists when a student, teacher, or course characteristic influences the evaluations made, either positively or negatively, but is unrelated to any criteria of good teaching, as much as increased student learning (p. 2)." Aleamoni (1999) summarized the literature of sixteen of the most common myths and biases assumed by faculty and administrators alike in regards to the validity of student ratings of instruction. “I take bias to mean something other than (or more than) the fact that student ratings may be influenced by conditions not under the teacher’s control or that conditions may somehow be ‘unfair’ to the instructor (Feldman, 2007 p. 95).” Typically studies involve correlating ratings with variables (student type, instructor type, course characteristics) to determine if any outside factors had an effect on student ratings of instruction. Overall, Aleamoni (1999) concluded that most myths were unproven and unsupported by research. Examples included: class size, relationship between expected grade and rating, whether students were taking a course for major/minor, and so forth. The most researched sources of bias included: gender, course type (required vs. elective, upper level vs. lower level) and class size.

The question of the effect of instructor gender on student ratings of instruction has been researched. Several studies have documented differences between male and female instructor student ratings of instruction, with mixed results. Centra and Gaubatz (2000) found that male students rated female instructors lower than their female classmates with no difference when the instructor was male. However, the effect size (magnitude of the difference, i.e. was it a relatively large difference) was modest at best. Goldberg and Callahan (1991) found that student ratings of instruction in business courses were dependent on gender. Throughout his extensive review of the literature, Feldman (1992) identified only a few instances where men received higher ratings than women. However Feldman found no cases where women received higher ratings than men. Basow and Montgomery (2005) found that female instructors were rated higher in both the natural sciences and humanities with no difference in engineering. Overall, Feldman (1992; 1993) concluded that there is a very weak relationship between instructor gender and student ratings of instruction ($r = .02$). Later studies (Feldman, 1992; Freeman, 1994; Hoffman & Oreopoulos, 2006; Wollert & West, 2000; Wright & Jenkins-Guarnieri, 2012) confirmed

Feldman's findings. However, there is some evidence that students rated same gender instructors higher than opposite gender instructors (Feldman, 1993; 2007), while Wright & Jenkins-Guarnieri (2012) found that the interaction between instructor gender and student gender did not impact student ratings of instruction.

When looking at course type and student ratings of instruction, several studies have been cited. The classic study is the meta-analysis conducted by Feldman (1978). Feldman reviewed studies from the previous thirty plus years related to the effect that student ratings of instruction have on course level. Feldman (1978) found that when other factors are controlled for (elective vs. required, academic motivation, and instructor characteristics, to name a few) the relationship is minimal at best. Scherr and Scherr (1990) concluded that students taking a course for an elective, as opposed to a requirement, found the subject matter more positive. Other researchers (Goldberg & Callahan, 1991; Hoyt & Lee, 2002a; 2002b; Kember & Leung, 2011; Wollert & West, 2000) have documented differences by subject/discipline area and student major (Gilmore et al, 1980; Oliver-Hoyo, 2008). However, when looking at ratings by subject matter (Cashin; 1990; Centra, 2003; 2009; Feldman, 1978; Santhanam & Hicks, 2002). The humanities and social sciences consistently scored higher than math, engineering, and physical science courses. Some reasons cited for the differences in disciplines included: (1) students have less developed quantitative skills (Cashin, 1990), (2) rapidly growing area of knowledge in the sciences (Centra, 2009), (3) the more sequential the course, the more difficult for students to succeed (Cashin et al, 1987; Cashin, 1990), and pressure and interest for faculty in the hard sciences, math, and engineering areas to produce research that may detract from their teaching (Centra, 2009). Although noteworthy, the issues presented above were not direct evidence of bias, but instead may be a difference in their effectiveness as instructors (Centra, 2009).

The potential bias of class size in student ratings of instruction has been well researched (Gilmore et al, 1980; Feldman, 1984, Marsh, 1987; Braskamp & Ory, 1994). Arias and Walker (2004) performed a quasi-experimental design in economics where the instructor taught the same way, gave the same lecture, exams, and grading policy. The only variable purported was class size. Students in smaller sections performed better on common final exam than larger sections. Likewise, in a comparison between two and four year institutions, Centra (2009) demonstrated that smaller class sections received higher ratings. This finding is both similar to and different from an earlier finding from Centra. In a study using a random sample of 10,000 classes, Centra & Creech (1976) found that small enrollment (under 35 students) and high (over 100 students) classes received the highest ratings and medium size enrollment courses scored the lowest ratings. Recently Guder et al (2009) analyzed the effect of increasing maximum class size by 50% in business courses. They found no difference in student ratings between the small and large class sizes.

Feldman (2007) summarized the issue of bias studies with the following quote,

> To put the matter (bias) in general terms, certain course characteristics and situational contexts-conditions that may not necessarily be under full control of the teachers-may indeed affect teaching effectiveness; and student ratings may then accurately reflect differences in teaching effectiveness (p. 98).

Even though the differences exist, and they might have an effect on teaching effectiveness, the threat to validity is minimal and student ratings of instruction should not be discounted. Centra (2009) using Student Instructional Report (SIR) data concluded that there was very little evidence of bias, but cautioned instructors and administrators to become aware of the potential for bias and the effects that bias could have in student ratings of instruction.

Experimenter variance validity studies are limited. These studies attempted to identify the effect that perceived knowledge of the instructor has on student ratings of instruction. The classic example here is the "Dr. Fox" lecture (Naftulin, Ware, & Donnelly, 1973).

The last type of validity studies are the conceptual framework studies of instruction. Conceptual framework studies attempted to identify common themes or elements of effective teaching by using advanced statistical methods like factor analysis. The more prominent studies (Kulik & Kulik, 1974; Feldman, 1976; Marsh, 1987; Marsh & Hocevar, 1984) have developed similar, yet different, results identifying effective traits of student instruction. What has been concluded is that effective teaching is multi-dimensional (Schmelkin et al, 1997). For example, an instructor may be organized, but lacks enthusiasm (Marsh, 1991).

Overall, student ratings of instruction have construct validity (Wright & Jenkins-Guarnieri, 2012), especially when institutions use nationally normed instruments such the IDEA and SEEQ forms. Administrators and faculty should become aware of potential threats to the validity, including potential biases. However, using documented techniques and forms will minimize those effects. Despite the present quantity of research on validity, faculty concerns over the usefulness and adequacy of using student ratings of instruction to measure teaching effectiveness still remain.

Faculty Concerns Over using Student Ratings of Instruction

One common criticism regarding student ratings of instruction is grading leniency. Faculty and instructors award high grades and/or lower academic standards in the anticipation of receiving higher student ratings (Culver, 2010; El Hassan, 2009; Huemer, 2005). Overbaugh's (1998) hypothesis was supported that demanding more from students in regards to learning and assignments would lead to lower ratings of instruction. Trout (2000), in his opinion piece to the

Academe magazine summed up his frustration with student ratings of instruction by stating, “For me, the key indictment against using the numerical forms to reward and punish the classroom behavior of instructors is that they encourage instructors to dumb down their teaching (pg. 2).” In essence, Trout felt that instructors would require less from their instruction by demanding less from students with regards to completing assignments, participating in class discussions, and knowledge and application for examinations.

Although a prominent belief in the research prior to the 1990s, most researchers now claim grading leniency is at best a slight factor in student ratings of instruction (Greenwald & Gillmore, 1997; Marsh, 2001). Centra (2003) found that when students identified courses as either too difficult, or too elementary, they were more likely to give unfavorable ratings. Courses that were identified as just right, in regard to rigor, received the highest marks. However, many faculty still believe that students give high ratings for lenient grading (Marsh & Roche, 1997), even though giving high grades will not alone produce high ratings (Centra, 2003; Greenwald & Gillmore, 1997; Marsh, 1987). Culver (2010) discovered that engagement, and engaging students in quality educational experiences, would have more of an effect on student ratings than leniency in grading, much to the delight of Kuh and his associates.

Goldman (1993) and Seldin (1993) argued against using student ratings of instruction as a stand-alone measurement of teaching effectiveness. Because of the perceived unreliability of student ratings of instruction, Goldman argued for a drastic overhaul of the faculty evaluation for teaching. Goldman proposed two alternative options: (1) a discipline specific student evaluation form developed by appropriate professional organizations and (2) a team of visiting colleagues that observe instruction and analyze syllabi. In both options, the results would be reported back to the faculty member’s home department and could be normalized and benchmarked against similar institutions and classes. Seldin (1993) took a holistic approach to the problem by incorporating portfolios, peer evaluations, and self-evaluations into the teaching assessment process. d'Apollonia & Abrami (1997) and El Hassan (2009) agreed that using the ratings are valid, but should be one piece of the evaluation process, not a stand-alone measure.

Most faculty understand the need for student ratings of instruction, but debate the usefulness of the data, and conclusions drawn from the ratings (Schmelkin et al, 1997; Trout, 2000). Bain (2004) conducted a study involving 63 instructors who survived a very selective interview process to be identified as the best in college teaching. Interviews, public presentations, syllabi, observational studies including videos, student ratings of instruction, and comments from colleagues were used to select the participants. Although the instructors were deemed high quality faculty in regard to teaching, even they disagreed about the reliability of student ratings of instruction and how they can be influenced by a number of factors, including grading leniency. In fact, one respondent stated his/her viewpoint on the usefulness of student ratings of instruction,

> High ratings from students indicate success only if I am satisfied with the quality that I am asking them to do intellectually, and that is reflected not in the ratings by in my syllabus, and the way I grade their work. Low ratings, on the other hand, usually tell me I've failed to reach my students (Bain, 2004, p. 166).

Comparisons by Academic Rank

Traditional American higher education has been defined by tenure status. Historically, faculty have identified themselves as either earning tenure or having achieved tenure. Until recently, with the growth of community/technical colleges and an increased emphasis on controlling costs for higher education, the tenure system has defined faculty classification in higher education. Full-time non-tenure track instructional faculty, part-time instructional faculty, supplemental or contingency faculty, and adjunct faculty have evolved into the traditional and non-traditional higher education system.

Full-time non-tenure track instructional faculty are individuals hired by academic departments on a recurring basis for the sole responsibility of instruction (University of Delaware, 2013). Conversely, tenure-track and tenured faculty have research and service responsibilities (University of Delaware, 2013). Institutions use various titles for these position types, but lecturer and instructor are two of the most commonly cited examples (University of Delaware, 2013).

In addition to full-time non-tenure track instructional faculty, institutions have been increasingly employing part-time instructional faculty, adjunct instructors, and supplemental instructors. The University of Delaware (2013) in conjunction with their Instructional Costs and Productivity Study, defined these persons as individuals paid to teach a course or courses out of the instructional budget with no recurring guarantee of contract. In some cases, these instructors provided instruction at no cost to the institution. For matters of consistency, these types of individuals are usually referred to as supplemental or contingent faculty.

Although tenure-track and tenured faculty still dominate the conversation on instruction in higher education, the evaluation and effectiveness of non-traditional faculty types have been studied. The central theme of the research involving non-traditional faculty focuses on the differences between tenure-track/tenured (TT/T) and non-traditional faculty on instructional quality and instructional effectiveness. In many comparisons, graduate teaching assistants (GTA) are included in the non-traditional group. Graduate teaching assistants are primarily graduate students pursuing a masters or doctoral degree in a specific instruction area that they are serving as the primary instructor. In some instances, GTAs serve as lab coordinators, discussion group leaders, or grading assistants.

Differences exist among non-traditional faculty. Bettinger and Long (2010) concluded that younger adjunct professors did better in academic subjects like math, English, history while older adjuncts scored better in professional courses such as law, nutrition, health sciences, and other applied practice courses. Other researchers have concluded that the instructor, or type of instructor, has a bigger effect on student ratings than the course type and level (Marsh & Overall, 1981).

Researchers have debated whether having non-traditional faculty teach entry level/survey courses jeopardizes student learning. Eagen and Jaeger (2008) found that full-time non-tenure track have no effect on student learning, the same finding was confirmed with GTAs teaching entry-level courses. Feldman (1987) found that time and effort devoted to research, generally the focus of tenure-track/tenured faculty, did not have a negative effect on teaching effectiveness. Conversely, contingent faculty, especially part-time faculty, tends to challenge students significantly less academically (Umbach, 2007). In addition, there was a negative outcome on student learning when supplemental and part-time faculty were used for entry level courses (Eagen & Jaeger, 2008).

Several studies have documented the differences in student ratings of instruction by faculty type and faculty rank. Braskamp et al (1984) found that traditional tenure-track/tenured faculty received higher ratings that graduate teaching assistants. Others findings include that overall, full-time non-tenure track faculty received the highest ratings of teaching effectiveness, while assistant professor (TT/T) received some of the lowest (Wollert & West, 2000). Student ratings of business instruction were dependent on instructor type, with adjunct faculty giving higher student grades (Goldberg & Callahan, 1991). Guder et al (2009) concluded that differences in student ratings of business instruction by faculty type did not differ among business students. Similar studies have shown no difference between adjunct/contingent faculty and traditional tenure-track/tenured faculty (Hellman, 1998; Hoffman & Oreopoulos, 2006; Landrum, 2009). Academic rank, and what types of courses are taught by what type of faculty, is one of the biggest differences among institution types. Specifically, at smaller liberal arts colleges and regional universities, it is common for tenure-track/tenured faculty to do a majority of the instruction while larger research institutions rely on the use of full-time non-tenure track, GTA, and supplemental faculty for instruction primarily at the lower levels (introductory and survey courses).

Course Level

Similar to instructor type, course level provides a systematic difference between courses and makes comparisons between lower and upper division or undergraduate and graduate courses difficult. Typically lower division courses are transactional and fact acquisition while upper division and graduate courses are discussion and knowledge discovery based. The traits that make a large introductory Psychology course instructor effective differ from a graduate seminar instructor. Commonalities of an effective instructor exist regardless of course level, Murray et al (1990) concluded that different traits and practices are needed for effective instruction and meaningful student learning to occur. In other words, what constitutes effective instruction in an entry level psychology course could be ineffective or less desirable for a graduate level seminar.

Recent studies related to course level concluded that elective courses received higher ratings on SRI than general education and lower division courses (Centra & Creech, 1976;

Santhanam & Hicks, 2002; Whitworth et al, 2002; Wollert & West, 2000; Zoller, 1992). Some of the earliest research, found little or no effect on course level/type when compared across different departments/institutions (Marsh & Overall, 1981; Scherr & Scherr, 1990).

Despite the general consensus about the effect of course level on SRI, the recent literature is attempting to discern the reason for these differences. Landrum (2009) found evidence among a number of social science departments that most lower division courses were taught by part-time instructors with little or no teaching experiences. Likewise, Nelson-Laird, Niskode-Dossett, and Kuh (2009) concluded via a qualitative study that faculty tend to think that general education/lower division course require less interaction and contact between faculty and students.

Regardless of the reason, there appears to be differences by course level/type and therefore more needs to be done to discover the reason for the differences. Some have argued that the student self-efficacy (Bandura, 2001) and the type of institution/department may account for differences in student learning. Therefore, having a clear understanding of student motivation, along with the SRI data, may help shed light on the differences between and among institutions regarding student learning.

Summary Related to Research of Student Ratings of Instruction

That being said, faculty and administrators seem to be warming to the idea of student ratings of instruction (Schmelkin et al, 1997). After all, as Seldin (1993) and others (Feldman, 2007; Huemer, 2005) have documented, the probability of using student ratings of instruction is only going to increase in the future. Therefore, educating faculty and administrators alike about the myths, biases, and use of student ratings of instruction would be prudent.

The reliability and validity of student ratings of instruction are generally considered to be both reliable and valid when using standardized (SEEQ, IDEA, etc.) rating forms by both researchers and practitioners (Spencer & Schmelkin, 2002). Although the literature supports this claim, many individuals (mostly faculty and administrators) still call for caution when using student ratings of instruction data (d'Apollonia & Abrami, 1997; Cheng & Marsh, 2010). Thus, the review and research of student ratings of instruction continue. However, most results in present studies are similar to the summary of the psychometric properties of student ratings of instruction offered up by Marsh (1987), "Student ratings are clearly multidimensional, quite reliable, reasonably valid, relatively uncontaminated by many variables often seen as sources of potential bias, and are seen to be useful by students, faculty and administrators (p. 369)."

Student Motivation

Student ratings of instruction (SRI) are an important piece of improving the student learning experience. Unfortunately SRI is not the only component of the student and classroom learning experience. The student experience, especially the undergraduate experience, encompasses many facets of the collegiate experience outside the classroom. Although efforts to improve student learning based on SRI data have been initiated, colleges and universities can take student motivation theory, specifically Albert Bandura's Social Cognitive Theory (Bandura, 1986; 1989; 1997; 2002), into account when addressing issues of student learning.

Although the literature is rich with motivational theories, most of them can be broken down into three main taxonomies: (1) humanistic, (2) behavioral, and (3) cognitive. To better understand the nuances of Social Cognitive Theory, a quick summary of the three motivational frameworks used to develop Social Cognitive Theory is needed.

Humanistic

The most famous humanistic theory of motivation is arguably Abraham Maslow's (1943, 1954) Hierarchy of Needs. The hierarchy is arranged in a pyramid structure with five needs, ordered from bottom to top: (1) physiological, (2) safety, (3) love, (4) esteem, and (5) self-actualization. The pyramidal structure is important because we must first satisfy our basic needs: physiological (food and water), safety, love, and esteem (self-esteem, respect, admiration from others), before self-actualization. Although the name implies a liner or hierarchal order, not all individuals feel the need for love at the same time. For some, esteem may be accomplished prior to love (Maslow, 1954). The main factor regarding basic needs is that they need to be met prior to the "growth" (self-actualization) stage of development. Self-actualization according to Maslow (1943, 1954) is when the individual begins to define their purpose in life, their potential. These needs, first physiological and ultimately purpose, are what drive human behavior.

Behaviorism

As Maslow is to humanistic motivational theory, B.F. Skinner (1953) and his operant conditional model are to behavioral motivation theorists. Skinner (1953) believed that all behavior, and, by extension, motivation was triggered by environmental cues and factors. Internal processes (cognitions) were independent from behavior. For example, during toilet training, every time a child uses the bathroom they are rewarded with encouragement and applause, a reward. Conversely, if a child does not use the bathroom and instead soils his or her clothes, he or she is talked to and informed that what he or she did is wrong, a punishment. Skinner believed that the use of rewards would reinforce positive behaviors while punishments, or the lack of positive reinforcement (rewards), would cause behavior to become weakened or extinguished. Even social-cognitive motivational theorists would agree that actions are affected by positive reinforcements (Locke et al, 1988).

Cognitive

Although the list of cognitive theorists is long, Jean Piaget (1964) and Lawrence Kohlberg's (1973) theories are considered the forerunners in cognitive theory. Both Piaget and Kohlberg believed personal development is comprised by how individuals processed observations and challenges they experienced, i.e. their cognitions, through a sequential series of stages (Kohlberg, 1973). In general, cognitive development psychologists view moral development as the product of one's individual basic structure for perceiving reality (Nichols & Day, 1982).

Modern theorists related to cognitive development, specifically Locke's (1968) Theory of Goal Commitment, are the most transferable to self-efficacy and Social Cognitive Theory. Commitment is the attachment or determination to reach a goal (Locke et al, 1988). Locke originally framed his theory around the relationship between goals and performance (Locke, 1968), Locke (1975) later refined the work to include the premise that personal choice regulates action. Locke and Latham's (1990; 2002) theory of Work Motivation and Satisfaction, specifically goal-setting, is important when talking about Social Cognitive Theory. According to Locke's theory, goals need to be: (1) specific, (2) measurable, (3) attainable, (4) realistic, and (5) time-bound, or SMART, for an acronym. Although goals need to be attainable, Locke (1968), and later confirmed by Latham and Yukl (1975), stated that specific goals increase performance, and when individuals take on appropriately challenging goals, the level of performance is higher for the individual.

According to Locke and Latham (1990; 2002) goals direct outcomes in four distinct ways: (1) goals are directive, i.e. linear, (2) goals are energizing and empowering, (3) persistence depends on the goals, and (4) goals affect our actions. Because goals are directive they lead individuals towards actions and learning activities in support of that goal. In other words, we focus on things that are important in achieving the goal and tune out matters that are not. In addition, difficult or ambitious goals lead to increased effort and persistence compared to easier or less ambitious goals. Lastly, depending on the goal, and the task management strategies conceived by the individual, the action taken by the individual, will vary. For example, for relatively modest goals, one can call upon personal history or examples to complete the task while more complex goals require a series of smaller tasks that may be undertaken in the immediate to build competence and map out a strategy to complete the overarching outcome. Task development is more likely to occur for individuals with high self-efficacy as opposed to those with lower levels of self-efficacy (Locke & Latham, 2002; Bandura & Wood, 1989).

Social Cognitive Theory

Goal setting theory leads nicely into Social Cognitive Theory precisely because of the importance of SMART goals on building self-efficacy and the importance of self-efficacy in task management. Albert Bandura went through several revisions, self-efficacy and Social Learning Theory, before finally settling on Social Cognitive Theory. Although most researchers focus on self-efficacy, Bandura's Social Cognitive Theory involves four processes that one uses in goal attainment. Goal attainment is the key to motivation according to Bandura, and motivation is best sustained by setting small goals that encompass a larger over-arching goal or achievement (Bandura, 1977; Locke & Latham, 1990). The four processes needed for goal attainment are: (1) self-observation, (2) self-evaluation, (3) self-reaction, and (4) self-efficacy. Although self-efficacy is the most researched and cited process in Bandura's model, the other three, cannot be ignored when discussing motivation related to self-efficacy. Bandura (1977; 1986; 1989) referred to this relationship as reciprocal determinism.

Reciprocal Determinism

Bandura defined reciprocal determinism as the interlocking effect between the environment, behavior, and the individual (Bandura, 1977). Bandura (1977; 1986; 1989) argued that the interaction between the environment, people and behavior, reciprocal determinism, is the basis for all human functioning. Drawing upon the humanistic, behavioral, and cognitive aspects of motivation, Bandura (1977) developed, and others (Bandura & Wood, 1989) would later refine, the Triadic Reciprocal Determinism Model to illustrate the relationship between the environment, an individual, and behavior.

Figure 1 *Triadic Reciprocal Determinism*

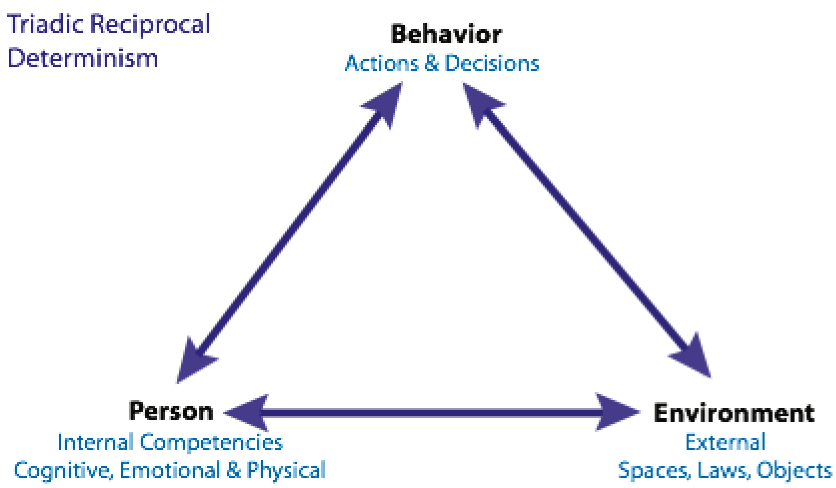


Figure copied from http://www.learningsolutionsmag.com/articles/474/overcoming-obstacles-to-avoid, accessed October 27, 2014.

Personal characteristics include the attributes, beliefs, values, goals, and emotions of the individual, cognitions (Bandura, 1978). Behavior includes the skills and actions that an individual performs, and the environment is the physical and social space that an individual is occupying, both directly and in-directly. As opposed to a strict behavioral approach, reinforcement and punishment affects the motivations of learners rather than directly causing behavior, in other words, learning cannot occur without cognitive processes (Bandura, 1989). Conversely, cognitive theorists feel that behavior is determined strictly by freewill, a directional determination.

An example of the relationship between the person, behavior, and the environment could be illustrated by a student who consistently sits in the back of a classroom. The student is known for being a loner and kind of shy. One day, one of the students in the front of the room invites the shy student to come sit up front. Over the course of the next few days, the previously shy student becomes a social butterfly. The student's behavior changed because of the reciprocal relationship between his environment and previous experiences.

Self-Observation

Within the context of reciprocal determinism, the first of the four learning processes, self-observation, involves discerning one's current behavior (Bandura, 2011). According to Social Cognitive Theory, taking a reflective stance on one's own behavior will lead to new sources of information moving forward to goal attainment. However, unless the self-observation is a regular occurrence in the life of the individual and it occurs relatively close to the event, the self-observation will do little to increase motivation towards goal attainment (Bandura, 2011).

Self-Evaluation

Self-evaluation is the assessment of performance in regards to the specific goal or outcome. Bandura (1977) stated that next to goal attainment, the most important cognitive source of motivation is self-evaluation. Goals need to be specific and valued for self-evaluation to lead to motivation. A goal of "doing good" on the next exam is difficult to assess, what constitutes "good?" Instead having a goal of 85% or better on the exam, or being in the top half (median value or above) of all scores in the class is more specific. Having normative and absolute standards for evaluation is important. For example, the results of a standardized test score could be both normative and absolute. The Graduate Record Examination (GRE) score of 600 on the verbal section of the exam is an absolute score. However, if that 600 is in the 80^{th} percentile of all scores during a specific testing year, the 80^{th} percentile is a relative score. Therefore, understanding the type of evaluation, either normative or absolute, an individual uses to assess performance is important. Otherwise the verbal score of 600 is meaningless if the individual goal was to be in the top ten percent of all test-takers.

Self-Reaction

Evaluation is good, but how the individual responds to what they learned from the evaluation contributes to learning and motivation. Goals that are met (or exceeded) push individuals to increase standards of performance (Bandura, 1989). Even the reaction to a less than desirable evaluation can be motivating. Bandura (1982; 1989; 1993) and others (Schunk, 1984, Chemers et al, 2001) found that individuals with high self-efficacy will see difficult situations or unexpected results as a motivating event. Conversely, individuals with low-self-efficacy would struggle or quit in the face of a difficult situation or failure.

Self-Efficacy

The cornerstone of Bandura's Social Cognitive Theory is the self-efficacy construct. Bandura (1977) defined self-efficacy as the belief structure that one holds about his or her ability to perform a specific task or series of tasks. Bandura (1997) later stated that self-efficacy as an internal question one must answer about his or her own ability before undertaking a task. The level of self-efficacy regarding a challenge or event will determine the choice he or she will make about the event, the level of commitment he or shew will give to the action, and the amount of discomfort they are willing to experience during the course of the event (Chemers et al, 2001). An individual with high self-efficacy regarding academic ability would set high goals (Locke and Latham, 2002) and would go to great strides to achieve even if it meant failure a time or two. A student with low self-efficacy regarding academic ability would avoid it, show little effort, or give up the first time it became difficult or challenging. Bandura (1977, 1982) identified four ways that individuals use to make a judgment regarding self-efficacy: (1) performance outcomes, (2) vicarious experiences, (3) verbal persuasion, and (4) physiological feedback.

Performance Outcomes

Performance outcomes, or past experiences, are the most important source of information regarding self-efficacy judgments (Bandura, 1977; 1982; Tschannen-Moran & Hoy, 2007). Outcomes that were successful lead to stronger views of self-efficacy. Likewise, negative outcomes lead to weaker views of self-efficacy. The earlier the negative outcome occurred in a sequence of events, the more damage to self-efficacy the event holds for the individual (Bandura, 1977). For example, struggling in basic algebra (poor test scores) leads to doubts about one's ability to understand higher order math operations than scoring poorly on a couple of tests in an advanced calculus class. The prior successes in algebra, trigonometry, and basic calculus motivates that student based on prior mastery experiences to continue working towards the goal of completing the course or even obtaining a degree that requires higher order math functioning.

Vicarious Experiences

Self-efficacy judgments are also nurtured based on how individuals perceive their abilities in relation to others (Bandura, 2011). Seeing someone succeed that you deem as an equal in a specific ability is a positive source of self-efficacy (Bandura, 2011). Conversely the opposite is true. Because of the influence others have on self-efficacy, Bandura (1982) stressed the importance of modeling. Modeling how to deal with challenging or frustrating outcomes in a positive manner leads to others concluding that in a similar situation (less than ideal outcome) they can overcome and learn from the unpleasant experience. Witnessing examples of people that quit or become discouraged, especially among peers, can be a damaging source of self-efficacy for individuals. Despite the importance of modeling the effects of modeling are not as strong and are vulnerable to change, because vicarious experiences are not as important as past experiences, (Bandura, 1977).

Verbal Persuasion

Self-efficacy is influenced by messaging. Through suggestion, people can be helped through an experience that has challenged them in the past (Bandura, 1977). However, the message bearer often carries more weight than the actual message. The more credible the message bearer, the more credible and effective the message (Bandura, 1997). However, the increased efficacy from messaging is often short lived (Bandura, 1982). Similarly, messaging that is in stark contrast to prior experiences is often ineffective due to the importance of personal experiences on self-efficacy (Bandura, 1977). Similar to vicarious experiences, the effects on self-efficacy are situational and of short duration.

Physiological Feedback

The last factor when making a judgment regarding self-efficacy is the physiological feedback from one's body. Feeling of nerves (sweaty palms, faintness) can immediately lower one's self-efficacy, regardless of how one felt prior to the experienced discomfort. In addition, the perceived discomfort can be damaging to self-efficacy. If one doubts their ability to perform a task (fear of failure), the perceived discomfort is often greater than the actual discomfort or difficulty (Bandura, 1977). The likelihood of success is higher when individuals are not surrounded by aversive arousal triggers (Bandura, 1982). Therefore, relaxation and de-sensitivity training are important coping mechanisms to limit the influence of physiological feedback on personal self-efficacy judgments (Bandura, 1977).

The Effects of Self-Efficacy

The basic principle behind self-efficacy is that people move toward activities that have high levels of self-efficacy and avoid activities with low levels of self-efficacy (Bandura, 1977, 1982). Once a determination regarding self-efficacy is made via one or all sources of information: past experiences, vicarious experiences, verbal persuasion, or physiological feedback, the expectation of the outcome depends on three dimensions: (1) magnitude, (2) generality, and (3) strength. Each of these dimensions have important implications regarding performance (Bandura, 1977). In addition, each of these dimensions can have minimal, moderate, or strong efficacy effects. Each of these dimensions is described in further detail below.

Magnitude

The magnitude of the expectation is the assumption one makes regarding the difficulty to complete a certain task or the maximum level of difficulty one can withstand in completing the task (Lunenburg, 2011). The more difficult the perceived task is the higher the magnitude of the effect (Bandura, 1977). An individual with a high self-efficacy regarding math ability would most likely consider the magnitude of a standardized math test as moderate at best.

Generality

Bandura (1977) argued effects on self-efficacy are determined by the scope of the event. Instances with very narrowly construed relevance offer very minimal self-efficacy expectations in a broad sense. Not surprisingly, the more transferable the mastery experience is for an individual, the stronger the effect has on the individual's self-efficacy as a whole. Lunenburg (2011) summarized generality as the extent that the experience is applicable to a diverse set of experiences. For example, mastering the perfect apple pie recipe has a very minimal generality expectation if it does nothing in adding to self-efficacy regarding an individual's ability to prepare a meal or bake other desserts.

Strength

Lastly the self-efficacy expectation is measured as minimal, moderate, or strong depending on the level of determination one demonstrates regarding seeing the task through to completion. It is a measure of resolve regarding individuals and how they deal with obstacles in their paths. Lunenburg (2011) defined strength related to self-efficacy as one's conviction to see the task through to completion. Individuals with strong expectations will overcome challenging experiences (Bandura, 1977; 2006). For example, individuals deemed strong regarding strength expectation would overcome poor academic advising and faculty mentoring in completing their undergraduate degree in biology, even if they were given poor or misleading advice during the duration of their coursework.

Limitations Regarding Self-Efficacy

The first limitation or warning regarding self-efficacy is the difference between self-esteem and self-efficacy (Linnenbrink & Pintrich, 2002). Self-esteem is the confidence in one's own worth or abilities in general (Linnenbrink & Pintrich, 2002). Whereas self-efficacy is situational and based on specific failures and accomplishments. Although the difference is slight, the two constructs are not interchangeable (Linnenbrink & Pintrich, 2002).

The second limitation, and is often confused when discussing self-efficacy, is relativity. Bandura (1997) warned about the danger of extrapolating efficacy beyond reasonable bounds. Because self-efficacy is based on self-perception, and not always achievement (Tschannen-Moran & Hoy, 2007), the possibility exists that an individual may assume his or her abilities are beyond their experiences (Bandura, 1997; 2006; Linnenbrink & Pintrich, 2002). Linnenbrink and Pintrich (2002) used this finding to warn that instructors should foster positive messaging, but realistic, self-efficacy beliefs in students. More specifically, it is important to encourage students but not beyond the student's current (or foreseeable future) ability.

Self-Efficacy and Student Learning

So why does self-efficacy matter related to student learning? In general, Graham and Weiner (1996) showed that self-efficacy was the number one predictor of behavior when compared to other student motivation theories. Bandura (2006) claimed self-efficacy is a central factor in people's lives. Even if you disagree on the centrality of self-efficacy, self-efficacy is one of the most important beliefs regarding student achievement (Linnenbrink & Pintrich, 2002). Chemers, Hu, & Garcia (2001) concluded that self-efficacy was directly related to performance and more importantly for this study, academic performance. Additionally, Margolis and McCabe (2003) concluded that self-efficacy is the key to improving the motivation of struggling students at the secondary level.

Bandura (1982) stated that self-efficacy influences learning and performance in three ways. First, students choose goals based on self-efficacy beliefs. Students with low self-efficacy choose lower goals for themselves compared to those with higher self-efficacy, limiting future opportunities. Secondly, self-efficacy affects effort. Students with lower self-efficacy do not exert as much effort in learning a new concept or task because they feel that ultimately they will fail. Lastly, self-efficacy influences persistence. Students who are challenged academically and have low self-efficacy are most likely to quit or avoid certain learning opportunities, while those with high levels of self-efficacy strive to complete a learning task even if they performed below expectations initially. Bandura (1977; 1986; 2002; Schunk, 1984; Locke and Latham, 1990) have argued that positive self-efficacy is critical for student learning, especially in a partially self-taught college or university environment.

Related to the higher education learning environment, how can instructors (regardless of type) raise the self-efficacy levels of their students and ultimately increase the level of student learning? Using employee motivation techniques, (Locke & Latham, 2002; White & Locke, 2000) and adapting it to a higher education environment, instructors could provide solid fundamental training, especially in entry level courses, to increase the probability of student success and build a base of mastery experiences in as many subject areas as possible. Secondly, instructors, especially at upper division and graduate courses, can model the academic mentality of intrigue and discovery for students by involving them in research and professionally related endeavors. Lastly, simple encouragement from the instructor can aid in self-efficacy, especially if the instructor is respected by the learner (Bandura, 1997).

Although one could argue that the overall motivational level of the small group or class affects teaching, motivation is not uniform across student level or even the class level. High self-efficacy, which produces higher levels of student motivation, does not overcome inferior understanding (Schunk, 1984) and therefore, SRIs are still the primary means to measure and improve teaching.

Addressing issues related to institution type (bachelor vs. doctoral granting), course level differences, and historically lower rated disciplines/subject areas (natural sciences vs. humanities) of SRI are still needed. In addition, the connection between students' ratings of instruction and student motivations need to be explored. If there is a connection between students' self-efficacy and measures of teaching effectiveness, a major pause should be given in the interpretation and implementation of policies based on the findings from student ratings of instruction.

Research Questions

As such, the following research questions will be addressed:

1. When looking at non-doctoral degree-granting institutions, is there an effect of course type, discipline, and course level on overall ratings of instruction?
2. When looking at doctoral degree-granting institutions, is there an effect of course type, discipline, and course level on overall ratings of instruction?
3. When looking at non-doctoral degree-granting institutions, is there an effect of course type, discipline, and course level on students' motivation to take the class?
4. When looking at doctoral degree-granting institutions, is there an effect of course type, discipline, and course level on students' motivation to take the class?
5. Is there a relationship between students' overall ratings of instruction and their motivation to take the class when course level is held constant?

Chapter 3 - METHODS

Introduction

Although student ratings of instruction and differences among student ratings of instruction have been examined, the issue of degrees awarded (bachelor vs. doctoral) remain relatively unexplored. In addition, the effect that course type, course level, and degrees awarded have on student motivation has not been studied. Chapter 3 discusses the quasi-experimental design, internal and external validity concerns, participants, reliability and validity of IDEA Center instruments, and methodology that were used in the present study

Quasi-Experimental Design

Because the experimenter did not have experimental control, a true experimental design was not possible. Kuehl (2000) described experimenter control as the actions a researcher takes to control the measurement/data collection technique, sampling method (randomization), choice of experimental design, and controlling for covariates. In a true experimental design for this study, the researcher would randomly assign students to various classes, give identical lesson plans, administer the SRI at the exact same point in the course sequence, and have equal numbers of students for each comparison of interest (institution type, course level, subject, gender of instructor, etc.).

A Quasi-Experimental Design, specifically a Static Group Comparison, was used to answer the proposed research questions. Campbell and Stanley (1963) defined a Static Group Quasi-Experimental Design as comparing a group that has experienced a treatment to a group that has not experienced a treatment with the intent to determine the effect of the treatment.

Validity concerns of the Experimental Design

The primary threats to internal validity in a Static Group comparison, according to Campbell and Stanley (1963), include maturation, mortality, selection, and the interaction of selection and mortality, while interaction of selection and x as the primary concern of external validity. Each of these threats is described in detail in the following section.

Maturation is the biological or psychological processes that change over time (Kuehl, 2000). In other words, the natural or learned differences an individual gains over a period of time is unrelated to the treatment of interest. For instance, do student ratings of instruction improve or worsen on factors unrelated to the instructor(s)? Mortality focuses on the decay or loss of participants over the course of a study (Campbell & Stanley, 1963). In this study, concerns over mortality were related to whether a differential number of students dropped out of a course or courses when compared to a different course. As opposed to true experimental design, there was no way to be certain that the groups (selection) were identical beyond the treatment received.

The last concern related to internal validity, interaction of selection and mortality, is the product of the two individual threats. Related to the current research study, the threat of the interaction of selection and mortality was that students with a high propensity of withdrawing from a course would pick a specific section or time of a course, thus changing the makeup of that course and affecting the overall rating of instruction.

Threats to external validity in a Static Group comparison are limited to the interaction of selection and x. The interaction of selection and x is concerned with the generalizability of the study. In some cases, the uniqueness of the experimental group and control group are so unique that the results are only transferable to the specific population of the study. The threat to external validity is greatly reduced when significant numbers exist for both the control and experimental group.

Addressing threats to validity is a problem in a non-randomized experimental design where subject-matching is not possible. The relative short-time frame (~16 weeks) for an academic term and the similarity of course types (Psychology 100 is relatively the same) across institutions address some of the concerns regarding internal validity threats, but they do not minimize the threats presented by the Static Group comparison. Threats to external validity in the current study are minimized due to the large sample size used in this study.

Participants

Due to the nature of this study, individual participants were not used. Instead of individual responses, class mean scores were used. Thus a participant is in actuality the class mean score for a particular statement of interest. For matters of clarity, and for ease of reading, the term "participants" will be used moving forward in lieu of "class mean score". Participants were students from 386,195 classes from the Individual Development and Educational Assessment (IDEA) Center database in Manhattan, Kansas. The IDEA Center is a non-profit organization designed to provide feedback and assessment to strengthen learning primarily through the collection of student ratings of instruction and administrator evaluations. Only student ratings of instruction from 2007 to 2011 were used for the present study. The average class size for the total population was 36.7 (median=22) students and a mode of 20 students, with a high of 3,535 students and a low of ten students.

After further examination of the population, 7,543 (2% of total population) classes were determined to be developmental (remedial), English as a Second Language (ESL) and First-Year Experience (FYE). Since these courses are not generally offered across all institutions of interest to this study, they were removed. From the remaining 378,652 classes, a frequency distribution revealed that 49,737 (12.9% of total population) classes had no "primary approach" provided for instruction. Considering the importance of class type on the study, these classes were removed. Further descriptive analysis revealed that 50,427 (13.1% of the total population) classes were reported as being as either a “combination” (serving multiple student types, upper class general education and upper class specialization for example) or had no student level indicated. Because of the importance of student level to the study, these classes were removed. Lastly, the researcher examined highest degree offered for the remaining, 278,488 classes. A frequency distribution revealed that a large number of classes were from two-year associate degree institutions (n=34,421), first-professional degree (n=5,660) institutions, and institutions that did not award a four-year degree (n=495). After removing these 40,576 (10.5% of the total population) classes this left a final total sample of 237,912 classes used for the study, representing 61.6% of the original population. The 237,912 classes originated from 372 different institutions in varying size and Carnegie classification. In interest of institutional anonymity, no state or regional identifiers were provided by the IDEA Center. Tables illustrating the final sample distribution can be found in Chapter 4.

Instrumentation

Two IDEA Center forms (Appendix A) were used for this study: (1) Student Ratings Diagnostic Form (SRDF) and (2) Faculty Information Form (FIF). The SRDF has two versions, the SRDF and the Short Form. The Short Form contains 12 statements relating to progress on various course requirements, specifically gaining factual knowledge, theory acquisition, teamwork, written communication skills, etc.

The items were presented in a Likert scale format from 1=No apparent progress, 2=Slight progress, 3= Moderate progress, 4= Substantial progress, 5= Exceptional progress. Six overall statements address student effort, academic background, and overall instructional quality. The responses were presented in Likert format with 1=Definitely false, 2=More false than true, 3=In between, 4=More true than false, and 5=Definitely true.

Two statements of particular interest were chosen related to this study. To capture a measure of student motivation, question 15, “I really wanted to take this course regardless of who taught it” was used. Likewise, to capture a measure of instruction, question 17, “Overall, I rate this instructor an excellent teacher” was used. Additionally, instructors were allowed to add twenty (20) course or institution specific questions that were not included in the IDEA Center database.

Most institutions used the paper-and-pencil method (80%) with the remainder using the online version. In addition to student responses, instructors provided some general information and descriptors about the class (i.e. teaching method, course requirements, distance learning, etc.).

The SRDF (See Appendix A) consisted of all components of the Short Form plus additional statements detailing instructor behavior, comparison to other courses taken at the same institution, and learning methods (i.e. technology, learning objectives, self-exploration). The SRDF consisted of 47 standardized items compared to 18 for the Short Form. Instructors were allowed to ask a maximum of 20 course or institution specific questions. The two questions of interest for this study were questions 39 and 41 respectively.

The Faculty Information Form (FIF) was completed by the instructor at a time that was distinct from the when students complete the SRDF. A full description of the FIF is found in Appendix A. The FIF asks demographic questions related to course type, level, instructor type, and course objectives. Examples of course objectives included factual knowledge, oral/written self-expression, understanding of intellectual/cultural activity, and teamwork. All twelve course objective items were given in a Likert scale format with 1=Minor or no importance, 2= Important, and 3=Essential. The twelve course objectives rated by the instructor on the FIF were rated by the student on the SRDF.

In addition to course objectives, instructors provided demographic information on eight factors: (1) class enrollment, (2) meeting time, (3) course type, (4) student type, (5) distance learning, (6) team taught, (7) course requirements, and (8) peripheral effects on learning. Course type is segregated by primary and secondary approach (primarily lecture with a secondary approach of studio for example). Student type differentiates by not only level (first year vs. graduate) but also degree requirement (general education vs. intended specialization). Distance learning and team-taught factors are yes/no forced response statements.

Course requirements consisted of a Likert scale response of 1=None (or little) required, 2=Some required, and 3=Much required to nine learning techniques ranging from writing, group work, and reading, to computer applications, oral communication, and creative/artistic/design endeavor. The peripheral effects statement consisted of a Likert scale response of 1=Positive effect on learning, 2=Neither a positive or negative effect, 3= Negative effect on learning, and 4=Cannot judge. Examples of peripheral learning include: physical facilities, instructor's desire to teach the course, students' level of effort to learn, and technical/instructional support.

Reliability of IDEA Center Instrumentation

The reliability of IDEA Center instruments were tested in an internal IDEA Center Technical Report Number 12 written by Hoyt and Lee (2002a). The split-half reliabilities were tested on all 47 items of the Student Ratings Diagnostic Form (SRDF). Only courses that had 13-17 respondents were used in the split-half comparison. A total of 44,447 classes from 1998-2001 were used in the study. The classes were randomly divided into two groups and means were calculated for each half. The Spearman-Brown Prophecy Formula was used to estimate the reliabilities for class averages of 12.5, 24.5, 42.5, and 60. These averages correspond to the IDEA Center class size categories of 10-14, 15-34, 35-49, and 50+. The results of the study are presented in Table 3.1.

Table 3.1 *SRDF Item Reliabilities: Split Half Reliabilities and Standard Errors*

	Reliability			Standard Error		
Class Size	Minimum	Maximum	Median	Minimum	Maximum	Median
10-14	0.39	0.90	0.79	0.21	0.34	0.27
15-34	0.56	0.95	0.88	0.16	0.26	0.20
35-49	0.69	0.97	0.93	0.13	0.21	0.16
50+	0.76	0.98	0.95	0.11	0.18	0.14

Figure 1: SRDF Item Reliabilities: Split Half Reliabilities and Standard Errors. Adapted from "IDEA Technical Report No. 12: Basic data for the revised IDEA system," by D.P. Hoyt, and E.J. Lee, 2002, p. 45-46.

Standard deviations were calculated to estimate the standard errors. The standard errors provided increased confidence in determining what the true mean score was for that variable.

Validity of IDEA Center Instrumentation

Validity can be defined as accurately measuring a specific trait or characteristic (Agresti & Finlay, 2009). The validity of IDEA Center instrumentation was focused on construct validity, and, specifically, convergent and divergent validity. Construct validity is defined as the ability to make conclusions based on the concept of interest (Kane, 2001). Specifically, based on the results of a particular study, one can conclude that you correctly identified the term of interest.

Convergent validity refers to the relationship between scores on different tests measuring the same variable (Cunningham et al, 2001). Issues of convergent validity related to this study were primarily focused on the relationship between overall instruction on the Student Ratings

Diagnostic Form and overall instruction on another standardized rating form such as the Student Evaluation of Educational Quality (SEEQ) form developed by Marsh (1982, 1984).

Divergent validity, sometimes referred to as discriminant validity, is the relationship between scores on different tests measuring two different variables (Holton et al, 2007). For an item/test to have high divergent validity, a relationship should be non-existent. An example of divergent validity for this study would be high scores on overall student ratings of instruction for the Student Ratings Diagnostic Form should not be similar to scores on a test measuring campus environment/climate.

Using class data (n=44,447) from 1998-2001, Hoyt and Lee (2002a) used four approaches to demonstrate the validity of IDEA Center ratings. The four approaches can be summarized as the: (1) correlational relationship between student progress and instructor's ratings of importance, (2) consistency of student ratings based on intuitive expectations, (3) differential validity of teaching method items, and (4) agreement between independent student and faculty ratings. A full description of Hoyt and Lee's (2002) validity findings is presented below.

To test the relationship between student progress ratings and instructor ratings of importance, a series of assumptions was needed. First, instruction is effective, instructors make meaningful and conscientious judgments when they rate the importance of an objective, and students make accurate rating of their progress. If these assumptions are true, then a positive correlation should exist. The results of the correlational comparisons are presented in Table 3.2.

Table 3.2 *Inter-Correlations of IDEA Faculty Information Form (FR) and IDEA Diagnostic Form (SR)*

	FR1	FR2	FR3	FR4	FR5	FR6	FR7	FR8	FR9	FR10	FR11	FR12
SR21	0.21											
SR22	0.14	0.17										
SR23	-0.04	-0.01	0.14									
SR24	0.00	-0.03	0.08	0.26								
SR25	-0.18	-0.14	0.10	0.15	0.39							
SR26	-0.32	-0.27	-0.04	0.10	0.17	0.37						
SR27	-0.18	-0.18	-0.11	-0.02	0.08	0.25	0.33					
SR28	-0.32	-0.26	-0.04	0.01	0.17	0.19	0.12	0.46				
SR29	-0.10	-0.10	0.08	0.12	0.12	0.05	-0.09	0.16	0.21			
SR30	-0.16	-0.11	0.03	0.05	0.13	0.08	0.02	0.15	0.08	0.28		
SR31	-0.21	-0.12	0.02	-0.02	0.08	0.08	0.03	0.23	0.07	0.16	0.27	
SR32	-0.09	-0.06	0.05	0.10	0.08	0.07	-0.02	0.06	0.06	0.11	0.08	0.09

Note: Coefficients represent correlations between student (SR21-32) and faculty ratings (FR-FR12) of the twelve learning objectives.
Figure 2: Inter-Correlations of IDEA Faculty Information Form (FR) and IDEA Diagnostic Form (SR). Adapted from "IDEA Technical Report No. 12: Basic data for the revised IDEA system," by D.P. Hoyt, and E.J. Lee, 2002, p. 6. (n=44,447).

The average correlations between instructor importance and student ratings of progress across all twelve objectives was +.265. Based on these findings, as well as an earlier validity study from 1973, student ratings on their progress have validity.

To demonstrate the consistency of student ratings of instruction with intuitive expectations, twenty teaching methods from the IDEA form were chosen because they were identified as desirable or potent teaching methods. If ratings are valid, there should be agreement between student progress and the frequency of these methods used by the instructor(s). Positive correlations between progress and frequency would provide evidence for validity. Only a handful of negative correlations existed with the highest magnitude being -.17.

The third approach, and the least conclusive approach, involved looking at the teaching methods that were most highly correlated to each course objective. In some cases, similar or identical lists of techniques were identified for different objectives. Table 3.3 illustrates the top eight objectives for each of the twelve teaching methods.

Table 3.3 *Relationship of Teaching Methods to Learning Objectives*

	Teaching Method											
	1	2	3	4	5	6	7	8	9	10	11	12
SR1			0.71									0.72
SR2	0.69	0.71	0.78	0.76	0.52	0.57	0.59	0.58	0.65	0.72	0.71	0.81
SR3												
SR4	0.73	0.72	0.79	0.79						0.70		0.72
SR5					0.77							
SR6	0.74	0.73	0.78	0.79			0.58			0.69		0.72
SR7			0.71	0.72		0.67	0.62	0.65	0.65		0.68	0.73
SR8	0.76	0.77	0.78	0.76	0.52	0.56	0.60	0.59	0.72	0.69	0.75	0.83
SR9									0.82			
SR10	0.70	0.69	0.71	0.71			0.58					
SR11										0.71		
SR12	0.69	0.68										
SR13	0.71	0.71	0.77	0.77		0.58	0.67	0.56	0.63	0.77	0.71	0.81
SR14					0.67				0.64			
SR15	0.66	0.68	0.79	0.78	0.59	0.66	0.59	0.62	0.73	0.71	0.69	0.80
SR16					0.53	0.59		0.68		0.75	0.72	
SR17												
SR18					0.67	0.57		0.60	0.63	0.69	0.66	0.74
SR19					0.53	0.78	0.61	0.76	0.68		0.71	
SR20									0.63			

Note: Coefficients represent the eight highest correlations for each teaching method for Medium (15-34) class size sections. Strongest three correlations for each teaching method are shaded in grey.
Figure 3: Relationship of Teaching Methods to Learning Objectives. Adapted from "IDEA Technical Report No. 12: Basic data for the revised IDEA system," by D.P. Hoyt, and E.J. Lee, 2002, p. 11-12.

For IDEA forms to be valid, and to independently measure objectives, lists should be somewhat dissimilar from each other.

The last approach, the agreement between independently obtained student and faculty ratings, is the approach most researched by Hoyt and Lee (2002a). If relationships between how an instructor and student rate a course exist, this finding would suggest the validity of student ratings. Although Hoyt and Lee did four studies to look at these relationships, and all four studies provided the validity of the IDEA form, the study that specifically dealt with motivation was chosen due to the nature of the current study.

In this study, student motivation for taking a course was examined, and how the motivation affects student ratings. In summary, as students progress through their academic program, and take classes more related to their major/career interest, student ratings of instruction increase. Based on these findings, the IDEA Center has initiated a series of adjustments to account for the differences between courses at different levels. If the adjustments were valid, they should be positive for general education and lower division courses and positive for upper division courses. Table 3.4 illustrates the adjustments made to IDEA form ratings.

Table 3.4 *Differences Between Adjusted and Unadjusted Ratings Among Five Types of Classes*

Criterion	Type of Class				
	General Education		Specialized/Major		
	Lower Division	Upper Division	Lower Division	Upper Division	Graduate/ Professional
21. Factual knowledge	+.08	+.01	-.06	-.07	-.06
22. Principles and theories	+.07	+.01	-.05	-.07	-.05
23. Applications	+.05	.00	-.04	-.08	-.11
24. Professional skills, viewpoints	+.05	+.01	-.03	-.04	-.08
25. Team skills	+.02	-.02	-.04	-.08	-.14
26. Creative capacities	+.06	.00	-.04	-1.0	-.14
27. Broad liberal education	+.06	-.01	-.07	-1.2	-.19
28. Communication skills	+.02	-.03	-.04	-.04	-.11
29. Find, use resources	+.06	+.02	-.02	-.05	-.08
30. Values development	+.06	.00	-.08	-.07	-.09
31 Critical analysis	+.02	-.01	-.04	-.06	-.09
32. Interest in learning	+.08	+.02	-.06	-.09	-.09
Excellent Teacher	+.04	.00	-.02	-.05	-.08
Excellent Course	+.11	+.06	-.08	-.08	-.12

Note: Figure 4: Differences Between Adjusted and Unadjusted Ratings Among Five Types of Classes. Adapted from "IDEA Technical Report No. 12: Basic data for the revised IDEA system," by D.P. Hoyt, and E.J. Lee, 2002, p. 52.

All adjustments were significant (P<.0001), and since the adjustments were in line with intuition, you can conclude the IDEA form is valid.

Statistical Procedures

A series of statistical procedures were used to answer the research hypotheses. Specifically, Analysis of Variance (ANOVA) techniques, correlation studies, and pairwise comparisons were the primary statistical techniques used in this study. Analysis of Variance is a statistical model used to measure an association between one or more predictor variables (categorical or ordinal) and a continuous (non-categorical) outcome variable (Agresti & Finlay, 2009).

First a Cronbach's Alpha was conducted on the dataset to test the internal consistency of the data from the IDEA center database. Cronbach's Alpha measures the average correlation of survey instrument items to test the reliability of the instrument (Santos, 1999). Generally, an alpha value of approximately .80 is needed for research applications (Streiner, 2003). Although specific statements of interest were questions 39 and 41 from the SRDF, all items on the SRDF were tested minus the statements related to learning objectives (statements 21-32) and any institutional specific questions asked (statements 49-67) by the institution. These class mean scores were not provided to the researcher for these statements. For the 35 remaining items on the SRDF, an overall alpha value of .972 was observed. The observed value is well above the threshold identified by Streiner (2003) for research applications. No statement, if deleted, would have improved the overall reliability of the instrument.

Besides determining if an association exits, researchers typically are concerned with the effect individual factors have on the outcome variable. Determining if main and/or interaction effects exist is an important process in any Analysis of Variance procedure. A main effect is described as the effect of a singular predictor variable on the outcome variable (Field, 2013). While an interaction effect is defined as the effect of predictor variable A on the dependent variable C is reliant on predictor variable B (Agresti & Finlay, 2009).

Pairwise comparison are generally of interest in a research study. Pairwise comparisons consist of comparing means from different treatments to see if a significant difference between the means exist (Kuehl, 2000). Specifically, are SRI from a business course higher than those from an education course? In this study, all pairwise comparisons used the Scheffé method. The Scheffé method is generally used for comparisons or contrasts suggested by the data, and are considered one of the most conservative pairwise comparison techniques with respect to type 1 error (Kuehl, 2000).

The last statistical technique to be used in this study was a correlation study. Correlation is a common technique used in statistical studies, and it is used to measure a linear relationship between two variables. The correlational relationship is usually measured by a Pearson product-moment correlation coefficient of r with a value of -1 to 1 (Agresti & Finlay, 2009). A strong positive relationship, i.e. hours spent studying and academic achievement, would have a correlation coefficient closer to 1. A strong negative relationship would have a correlation coefficient closer to -1. A coefficient value near zero suggests no relationship exists.

Limitation of the Study

There are a few noteworthy limitations regarding this study. First, there was a lack of experimenter control. Since the data were generated from a secondary data source (IDEA Center diagnostic forms), all the research questions were dependent on what could be extracted from the responses on the diagnostic forms. Questions regarding institutional teaching differences, and other institutional factors not captured in the IDEA Center database, should be studied at a later date.

Secondly, since the data are limited to the IDEA Center database, and not all institutions use the IDEA Center diagnostic forms, generalizability beyond the population of IDEA Center participating institutions was difficult. However, given the geographic diversity and number of classes incorporated in the IDEA Center database, one could argue for the transferability of these findings to other institutions.

Much has been written about the accuracy of SRI. Although the research contained both positive and negative recommendations about the use of SRI, the consensus is that SRI is valid and reliable (Centra 2003, Marsh, 1984,). Despite the general consensus regarding the validity and reliability of SRI, apprehension exists among faculty and administrators alike about the apprehension in using SRI to draw conclusions.

Lastly, there were no standardized administration format or time. Specifically, students could fill out the instructor evaluation at various points throughout the academic term. The IDEA Center asked that you do not fill out the Short Form/SRDF before the halfway point of the course or on the last day the class meets. Beyond those two requirements, when the Short Form/SRDF is administered is up to the individual institution. Therefore, comparing student learning between two sections of an identical course (Psychology 101 for example) may be difficult since the collection points may differ. If Instructor A chose to collect feedback at the 75% point in the class while Instructor B chose the 90% point, it would be reasonable to conclude that Instructor B's section would have more "material" to base their conclusions. On the other hand, collecting data earlier in the semester may account for individuals that are struggling academically and may not be present later in the semester, i.e. students who end up dropping the class. Therefore, the variance between administration dates may differ widely.

Summary

Based on the work of Hoyt and Lee (2002a; 2002b) the reliability and validity of IDEA SRI are generally accepted as reliable and valid. Based on the validity studies conducted, IDEA forms demonstrate construct validity and provide generalizable results to the general academic community as a whole. Chapter 4 details the results of this study based on the research questions and research design presented earlier.

Chapter 4 - RESULTS

Introduction

This chapter provides results of the statistical analyses used to test the research questions. The results were organized around the research questions provided in Chapter 2. Descriptive statistics for the four independent variables were also provided. For each independent variable, a frequency table with appropriate central tendency statistics was provided in the appendices. To test research questions one through four, a series of univariate analyses of variance (ANOVA) were conducted. A partial correlation between overall ratings of instruction and student motivation was conducted to test the last research question.

Descriptive Statistics for Independent Variables

The four independent variables of interest for this study included: (1) degree granting institution (two levels), (2) course type (three levels), (3) discipline (five levels), and (4) student type (five levels). Descriptive statistics from the final sample are provided below.

Degree Granting Type of Institution

The IDEA Center provided 19 levels for highest degree offered at an institution. As described earlier, the only levels of interest for this study included: Baccalaureate, Masters, Beyond Masters but Less than Doctorate, and Doctorate. The counts from the final sample are presented in Table 4.1.

Table 4.1 *Highest Degree Offered: IDEA Center Classifications*

	Frequency	Percent
Baccalaureate	27,716	11.6%
Masters	64,390	27.1%
Beyond Masters But Less than Doctorate	8,776	3.7%
Doctorate	137,030	57.6%
Total	237,912	

Since this study was limited to differences between doctoral and non-doctoral institutions, the Baccalaureate, Masters, and Beyond Masters but Less than Doctorate were collapsed into a new categorical variable labeled "non-doctoral". It represented 100,882 cases or 42.4% of the sample. Doctorate institutions were renamed as "Doctoral" representing 137,030 cases or 57.6% of the sample.

Course Type

The Faculty Information Form (FIF) allowed the course instructor(s) to choose one of ten possible "primary teaching methods" in question one of the form (See Appendix A). As discussed in Chapter 3, classes identified as "Other" were removed from the final sample. The final distribution of primary teaching methods are presented in Table 4.2.

Table 4.2 *Primary Teaching Method: Classifications from Faculty Information Form (FIF)*

	Frequency	Percent
Lecture	140,954	59.2%
Discussion	32,653	13.7%
Seminar	14,284	6.0%
Skill/Activity	24,732	10.4%
Lab	12,735	5.4%
Field Experience	1,926	0.8%
Studio	6,259	2.6%
Multi-Media	3,146	1.3%
Practicum/Clinic	1,223	0.5%
Total	237,912	

For ease of comparison, and to align with other studies documented in Chapter 2, Lecture, Discussion, and Seminar were combined into one category (n=187,891 or 79% of sample), Lab and Studio were combined into another category (n=18,994 or 8% of sample), and the remaining teaching methods were combined into the last category (n=31,027 or 13% of sample).

Discipline

The Faculty Information Form (FIF) allowed the course instructor(s) to choose a discipline code, modified from the Department of Education Classification of Instructional Program (CIP) codes that best reflected the subject area for the class being rated (See Appendix B). Additional discipline codes were provided on the IDEA Center website and could be used if one of the subject areas on the FIF did not correspond to the subject area.

As previously mentioned, classes identified as First-Year Experience, Developmental, English as a Second Language were removed from the final sample. Based on the literature review, four distinct disciplines/subject areas were used for the comparison: Liberal Arts; Education; Business; and Science, Technology, Engineering, and Mathematics (STEM).

Considering the breadth of subject areas used in the IDEA Center Database, and for the sake of larger cell sizes, IDEA Center discipline codes were combined as follows. Liberal Arts were defined using discipline code 2400 (Liberal Arts & Sciences, General Studies and Humanities). Education classes were defined by using all discipline codes within the range of 1300 to 1332, excluding discipline codes 1301 and 1327. No class descriptors were provided for these codes to insure they were Education related classes. Business classes were defined by using all discipline codes within the range of 5200 to 5216. A partial listing of discipline codes can be found in Appendix B. A full listing can be found on the IDEA Center website by searching for "Discipline/Department Codes".

Considering the popularity of STEM research, there were a host of definitions used to define what did or did not constitute a STEM subject area. Based on the literature review, and the experience of the researcher, the CIP code listing of STEM subject areas as defined by the Department of Homeland Security (2014) was used as the basis for comparison. A full listing of STEM CIP codes can be found in Appendix B. The only manipulation to the STEM listing consisted of including discipline code 1305 (CIP code 13.0501) and discipline code 1306 (CIP code 13.0601 and 13.0603) in the Education group and discipline code 5213 (CIP codes 52.1301, 52.1302, 52.1304, and 52.1399) in the Business group. Including these discipline codes in the Education and Business groupings accounted for 435 (2.1% of group total) and 857 (3.3% of group total) classes remaining in each group respectively.

Classes not falling into one of the four disciplines of interest were classified as "Other" and constituted 55.9% of the sample. Because of the number of classes classified as "Other," it was determined to keep "Other" classes in the final sample. Examples of "Other" included: Consumer Sciences, Theological and Ministerial Studies, Psychology, and various health professions to name a few. Although this was a high percentage of classes outside the four disciplines of interest, considering the range of classifications provided by the Department of Education, and the need for the IDEA Center to align as closely as possible with external classification systems, this outcome seemed reasonable. Additionally, removing these classes would further decrease the sample from 61.6% of the original population to 27.2% of the original population while making some cell counts smaller than their current levels. The final distribution of modified disciplines can be found in Table 4.3.

Table 4.3 *Discipline Counts as Defined by Researcher*

	Frequency	Percent
STEM	50,839	21.4%
Education	20,777	8.7%
Business	25,859	10.9%
Liberal Arts	7,454	3.1%
Other	132,983	55.9%
Total	237,912	

Student Type

The Faculty Information Form (FIF) allowed the course instructor(s) to choose one of six possible "principle type of students" in question five of the form (See Appendix A). Those classes identified as teaching to "Combination" and having no value recorded for student type were removed from the final sample. The final distribution for student type is presented in Table 4.4.

Table 4.4 *Principal Type of Student Enrolling in This Course*

	Frequency	Percent
Lower Level, General Education	74,136	31.2%
Lower Level, Specialization	48,660	20.5%
Upper Level, General Education	19,887	8.4%
Upper Level, Specialization	70,922	29.8%
Graduate/Professional	24,307	10.2%
Total	237,912	

Lower Level can be defined as first-year or sophomore level courses generally at the introductory or survey level. Upper Level can be defined as beyond sophomore level courses generally focused on a narrower subject area but more in-depth than the lower level course.

Cross Tabulation by Degree Granting Type of Institution

Because the research questions were framed around degree granting type of institution, a cross tabulation by degree-granting type of institution and course type, discipline, and student type is presented in Table 4.5.

Table 4.5 *Cross Tabulation by Degree Granting Type of Institution*

	Non-doctoral		Doctoral	
	Frequency	%	Frequency	%
Course Type				
Lower Level, General Education	34,771	34.5%	39,365	28.7%
Lower Level, Specialization	23,913	23.7%	24,747	18.1%
Upper Level, General Education	9,172	9.1%	10,715	7.8%
Upper Level, Specialization	28,060	27.8%	42,862	31.3%
Graduate/Professional	4,966	4.9%	19,341	14.1%
Total	100,882		137,030	
Discipline				
STEM	17,826	17.7%	33,013	24.1%
Education	6,639	6.6%	14,138	10.3%
Business	10,739	10.6%	15,120	11.0%
Liberal Arts	5,102	5.1%	2,352	1.7%
Other	60,576	60.0%	72,407	52.8%
Total	100,882		137,030	
Student Type				
Lecture/ Discussion/ Seminar	78,238	77.6%	109,653	80.0%
Laboratory & Studio	7,615	7.5%	11,379	8.3%
Activity/ Field/ Media/ Practicum	15,029	14.9%	15,998	11.7%
Total	100,882		137,030	

Research Question 1

A three-way analysis of variance (ANOVA) was conducted to evaluate the relationship between overall student ratings of instruction and course type, discipline, and student type at non-doctoral granting institutions. The means and standard deviations for overall ratings of instruction for course type, discipline and student type are presented in Appendix C. The results of the three-way ANOVA are reported in Table 4.6.

Table 4.6 *Three-Way Analysis of Variance Output for Teaching Effectiveness at non-doctoral Institutions*

Dependent Variable: Overall I rate this instructor an excellent teacher (non-doctoral)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	557.159a	73	7.632	21.994	.000
Intercept	28112.055	1	28112.055	81010.738	0.000
Discipline	28.470	4	7.118	20.511	.000
Course Type	3.576	2	1.788	5.153	.006
Student Type	3.097	4	.774	2.231	.063
Discipline*Course Type	4.647	8	.581	1.674	.099
Discipline*Student Type	9.539	16	.596	1.718	.036
Course Type*Student Type	6.507	8	.813	2.344	.016
Discipline*Course Type*Student Type	20.517	31	.662	1.907	.002
Error	34981.336	100806	.347		
Total	1875719.705	100880			
Corrected Total	35538.495	100879			

a. R Squared = .016 (Adjusted R Squared = .015)

Note: No class mean score for teaching effectiveness were provided by the IDEA Center for 2 classes. Therefore, total sample size for the model is 100,880 compared to 100,882 displayed in Table 4.5.

The highest order interaction term of discipline, course type, and student type on teaching effectiveness was significant (F (31, 100,806) = 1.907, p = .002, η^2 = .001). Although the highest order interaction term was significant, it is important, on average, to report the main effects. Students reported significantly different levels of teaching effectiveness by discipline (F (4, 100,806) = 20.511, p = <.001, η^2 = .001) at non-doctoral institutions. In addition, students reported significantly different levels of teaching effectiveness by course type (F (2, 100,806) = 5.153, p = .006, η^2 = .000). However, students showed no difference in teaching effectiveness across student type (F (4, 100,806) = 2.231, p = .063, η^2 = .000). When examining effect size (η^2) for each of the independent variables (discipline, course type, and student type), as well as all four interaction terms, and using the cut-off values provided in Cohen (1992) as a guide to determine the strength of the effect size, no notable effect size was observed ($r < .10$) from the factors in the model. In other words, no one independent variable, or interaction term, accounted for more than 1% of the total variance among teaching effectiveness.

The three-way interaction effect of course type, discipline, and student type was found to be statistically significant. This finding indicated that the pattern of differences between course type, discipline, and student type was significantly altered when all three factors were considered on non-doctoral granting institutions. After examining the various mean score plots, the interaction effect seemed the strongest when looking at course type by discipline and student type. Although Lecture/Discussion/Seminar mean values remained relatively consistent in relation to other Lecture/Discussion/Seminar mean values by discipline over student type, Laboratory & Studio and Activity/Field/Media/Practicum values were the least consistent. Business and STEM disciplines showed the greatest interaction effect on mean values. Figures 2 and 3 provide an illustration of the interaction effect between course type and discipline when holding student type constant.

Figure 2 *Non-Doctoral: Lower Level General Education Teaching Effectiveness*

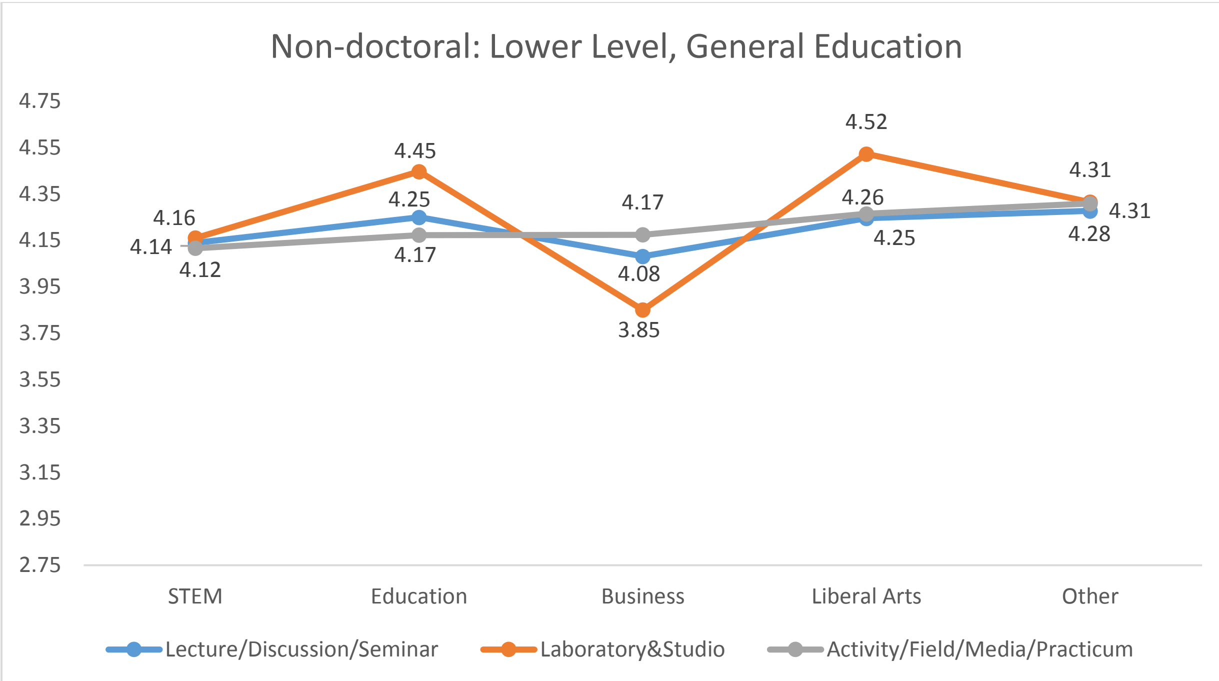
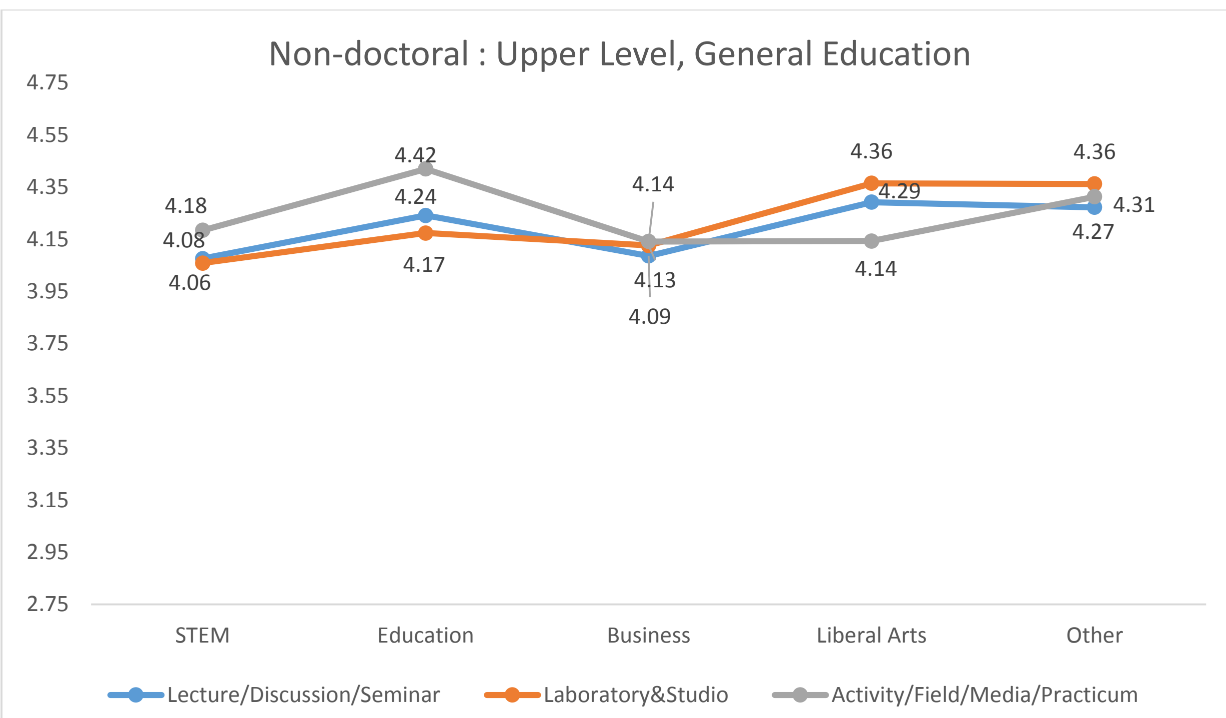


Figure 3 *Non-Doctoral: Lower Level General Education Teaching Effectiveness*



The mean ratings for STEM, Education, Business, and Liberal Arts determined by combining course type and student type were 4.17 (SE=.02), 4.32 (SE=.03), 4.15 (SE=.02), and 4.33 (SE=.05) respectively. Using the Scheffé statistic, STEM disciplines were rated lower than all other disciplines while Education was rated higher than other disciplines. The mean ratings for Lecture/Discussion/Seminar, Laboratory & Studio, and Activity/Field/Media/Practicum determined by combining discipline and student type were 4.27 (SE=.01), 4.29 (SE=.03), and 4.21 (SE=.02) respectively. Lecture/Discussion/Seminar received the lowest ratings while and Laboratory & Studio were rated highest. There was no significant main effect found for student type.

Research Question 2

A three-way analysis of variance (ANOVA) was conducted to evaluate the relationship between overall student ratings of instruction and course type, discipline, and student type at doctoral granting institutions. The means and standard deviations for overall ratings of instruction for course type, discipline and student type are presented in Appendix C. The results of the three-way ANOVA are reported in Table 4.7.

Table 4.7 *Three-Way Analysis of Variance Output for Teaching Effectiveness at Doctoral Institutions*

Dependent Variable: Overall I rate this instructor an excellent teacher (Doctoral)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1167.464a	72	16.215	45.199	.000
Intercept	23970.466	1	23970.466	66818.660	0.000
Discipline	183.597	4	45.899	127.946	.000
Course Type	4.830	2	2.415	6.732	.001
Student Type	6.700	4	1.675	4.669	.001
Discipline*Course Type	29.587	8	3.698	10.309	.000
Discipline*Student Type	26.530	16	1.658	4.622	.000
Course Type*Student Type	4.676	8	.584	1.629	.111
Discipline*Course Type*Student Type	52.449	30	1.748	4.873	.000
Error	49087.348	136833	.359		
Total	2510845.854	136906			
Corrected Total	50254.812	136905			

a. R Squared = .023 (Adjusted R Squared = .023)

Note: No class mean score for teaching effectiveness were provided by the IDEA Center for 124 classes. Therefore, total sample size for the model is 136,906 compared to 137,030 displayed in Table 4.5.

The highest order interaction term of discipline, course type, and student type on teaching effectiveness was also significant (F (30, 136,833) = 1.748, p < .001, η^2 = .001). Although the highest order interaction term was significant, it is important, on average, to report the main effects. Students reported significantly different levels of teaching effectiveness by discipline (F (4, 136,833) = 127.946, p = <.001, η^2 = .004) at doctoral institutions. In addition, students reported significantly different levels of teaching effectiveness by course type (F (2, 136,833) = 6.732, p = .001, η^2 = .000). Finally, a significant difference in teaching effectiveness amongst student type (F (4, 136,833) = 4.669, p = .001, η^2 = .000) existed. When examining effect size (η^2) for each of the independent variables (discipline, course type, and student type), as well as all four interaction terms, and using the cut-off values provided in Cohen (1992) as a guide to determine the strength of the effect size, no notable effect size was observed (r < .10) from the factors in the model. In other words, no one independent variable, or interaction term, accounted for more than 1% of the total variance among teaching effectiveness.

The three-way interaction effect of course type, discipline, and student type was found to be statistically significant. This finding indicated that the pattern of differences, noted above, between course type, discipline, and student type was significantly altered when all three factors were considered on doctoral granting institutions. After examining the various mean score plots, the interaction effect seemed the strongest at course type by student type and discipline. For all disciplines, Lecture/Discussion/Seminar mean scores were relatively flat, outside of STEM where the mean values trended upwards, especially among Upper Level Specialization and Graduate/Professional students. Conversely, the other two course types showed significant variability across all five disciplines. With the exception of Business and STEM disciplines having lower mean values for Laboratory & Studio courses at the Upper Level General Education level, no clear pattern exists. Figures 4 and 5 provide an illustration of the interaction effect between course type and discipline when holding student type constant.

Figure 4 *Doctoral: Lower Level General Education Teaching Effectiveness*

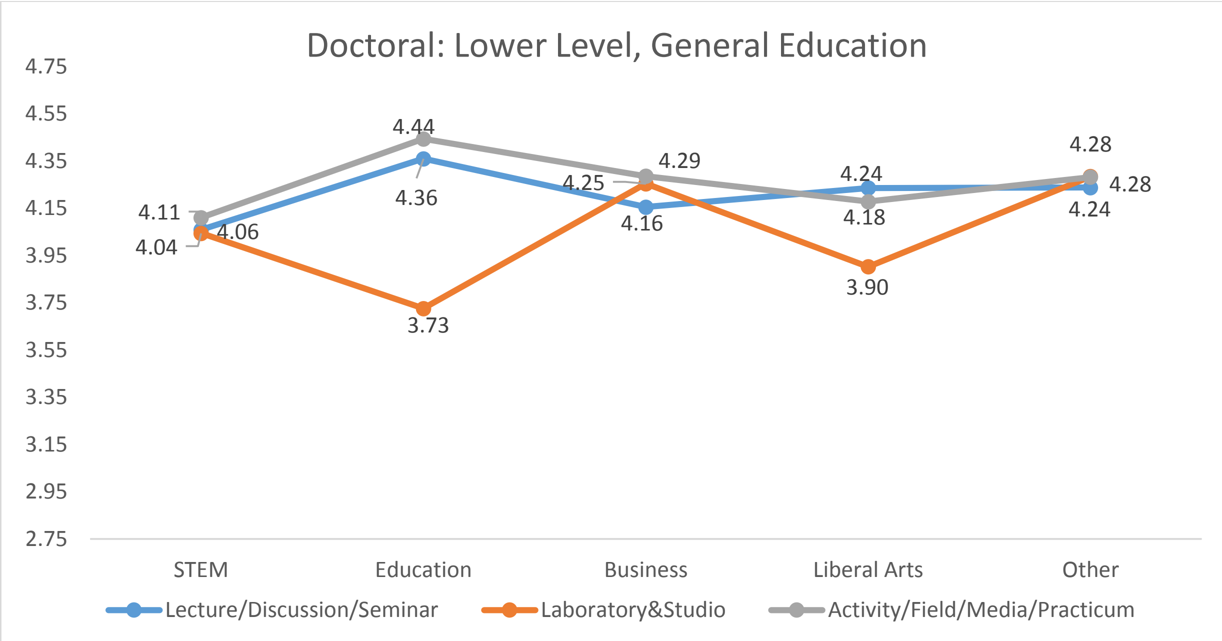
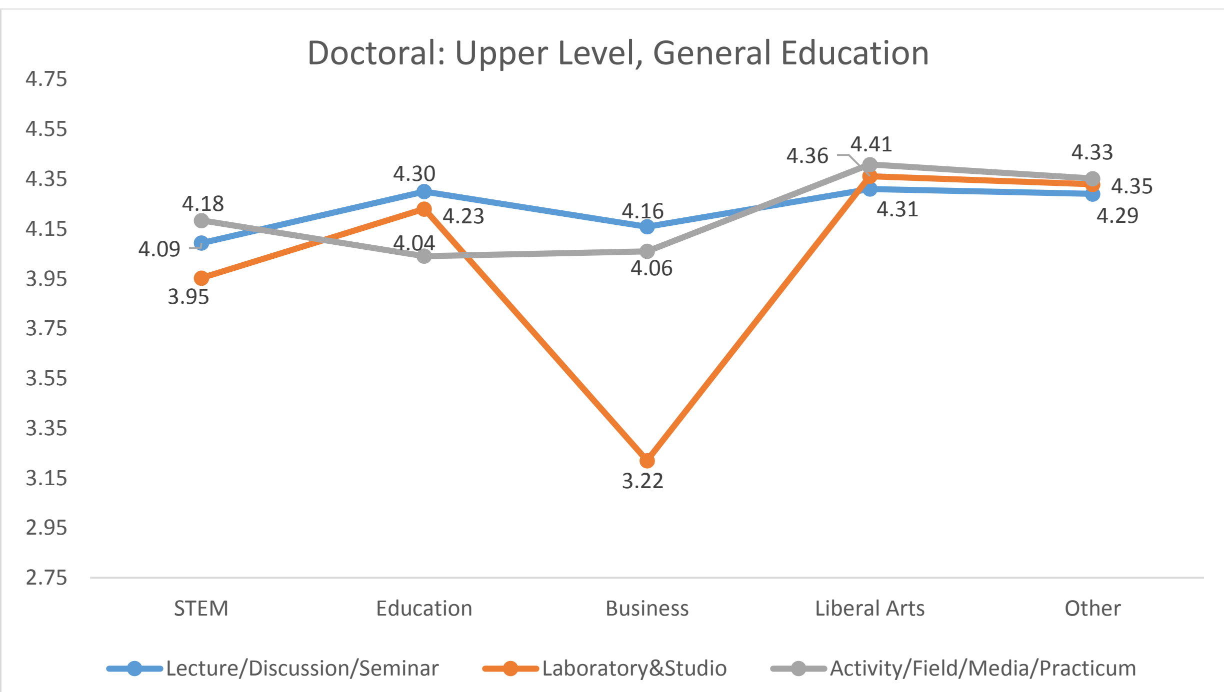


Figure 5 *Doctoral: Upper Level General Education Teaching Effectiveness*



The mean ratings for STEM, Education, Business, and Liberal Arts determined by combining course type and student type were 4.11 (SE=.01), 4.27 (SE=.02), 4.07 (SE=.02), and 4.37 (SE=.06) respectively. Using the Scheffé statistic, STEM disciplines received lower ratings than all other disciplines. No significant difference between Liberal Arts and Other existed. The mean ratings for Lecture/Discussion/Seminar, Laboratory & Studio, and Activity/Field/Media/Practicum determined by combining discipline and student type were 4.26 (SE=.01), 4.12 (SE=.03), and 4.27 (SE=.03) respectively. Laboratory and Studio course types were rated lower than both Lecture/Discussion/Seminar and Activity/Field/Media/Practicum. Conversely, Activity/Field/Media/Practicum course types were rated higher than the other two types. The mean ratings for lower level general education, lower level specialization, upper level general education, upper level specialization, and graduate determined by combining course type and discipline were 4.17 (SE=.02), 4.23 (SE=.03), 4.15 (SE=.03), 4.27 (SE=.02), and 4.29 (SE=.05) respectively. In general, as students progressed throughout their coursework, their mean ratings of instruction score increased.

Research Question 3

A three-way analysis of variance (ANOVA) was conducted to evaluate the relationship between overall student motivation and course type, discipline, and student type at non-doctoral granting institutions. The means and standard deviations for student motivation for course type, discipline and student type are presented in Appendix C. The results of the three-way ANOVA are reported in Table 4.8.

Table 4.8 *Three-Way Analysis of Variance Output for Student Motivation at non-doctoral Institutions*

Dependent Variable: I really wanted to take this course regardless of who taught it (non-doctoral)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5150.414a	73	70.554	276.128	.000
Intercept	18005.356	1	18005.356	70468.283	0.000
Discipline	67.745	4	16.936	66.284	.000
Course Type	8.057	2	4.028	15.766	.000
Student Type	80.830	4	20.207	79.087	.000
Discipline*Course Type	33.983	8	4.248	16.625	.000
Discipline*Student Type	34.307	16	2.144	8.392	.000
Course Type*Student Type	14.491	8	1.811	7.089	.000
Discipline*Course Type*Student Type	42.387	31	1.367	5.351	.000
Error	25756.949	100806	.256		
Total	1176588.952	100880			
Corrected Total	30907.363	100879			

a. R Squared = .167 (Adjusted R Squared = .166)

Note: No class mean score for motivation were provided by the IDEA Center for 2 classes. Therefore, total sample size for the model is 100,880 compared to 100,882 displayed in Table 4.5.

The highest order interaction term of discipline, course type, and student type on student motivation was also significant (F (31, 100,806) = 5.351, p = <.001, η^2 = .002). Although the highest order interaction term was significant, it is important, on average, to report the main effects. Students reported significantly different levels of student motivation by discipline (F (4, 100,806) = 66.284, p = <.001, η^2 = .003) at non-doctoral institutions. In addition, students reported significantly different levels of motivation by course type (F (2, 100,806) = 15.766, p = .000, η^2 = .000). Finally, a significant difference for student motivation amongst student type (F (4, 100,806) = 79.087, p = <.001, η^2 = .003) existed. When examining effect size (η^2) for each of the independent variables (discipline, course type, and student type), as well as all four interaction terms, and using the cut-off values provided in Cohen (1992) as a guide to determine the strength of the effect size, no notable effect size was observed (r < .10) from the factors in the model. In other words, no single independent variable, or interaction term, accounted for more than 1% of the total variance among student motivation.

The three-way interaction effect of course type, discipline, and student type was found to be statistically significant. This finding indicated that the pattern of differences, noted above, between course type, discipline, and student type was significantly altered when all three factors were considered on non-doctoral granting institutions. After examining the various mean score plots, the interaction effect seems strongest at course type by student type and discipline. For all disciplines, Lecture/Discussion/Seminar mean scores were relatively consistent with specialization and Graduate/Professional mean values higher than general education. In opposition, the other two course types showed significant variability, or lack thereof, dependent on discipline. For example, STEM mean values of motivation were nearly identical across all student types, while Education mean values of motivation varied greatly by course type and student level. For the Other discipline, mean values of motivation showed very little interaction outside of Graduate/Professional students. Figures 6 and 7 provide an illustration of the interaction effect between course type and discipline when holding student type constant.

Figure 6 *Non-Doctoral: Lower Level General Education Student Motivation*

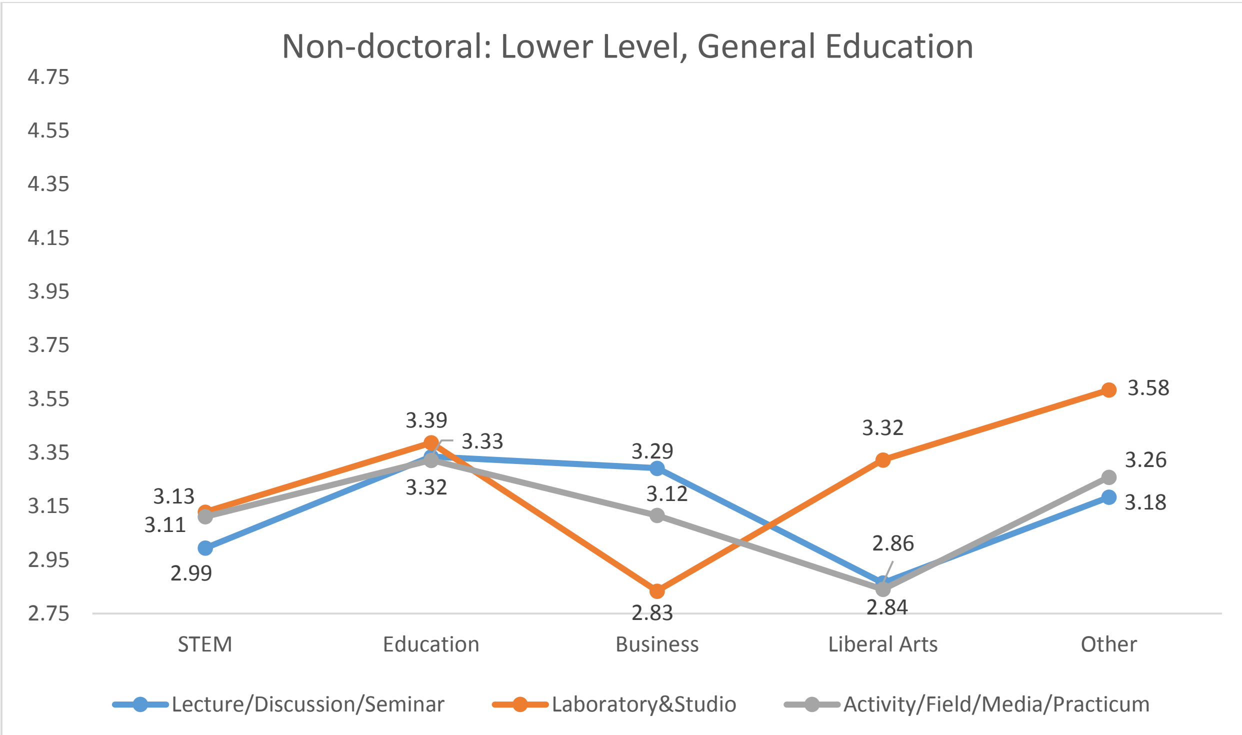
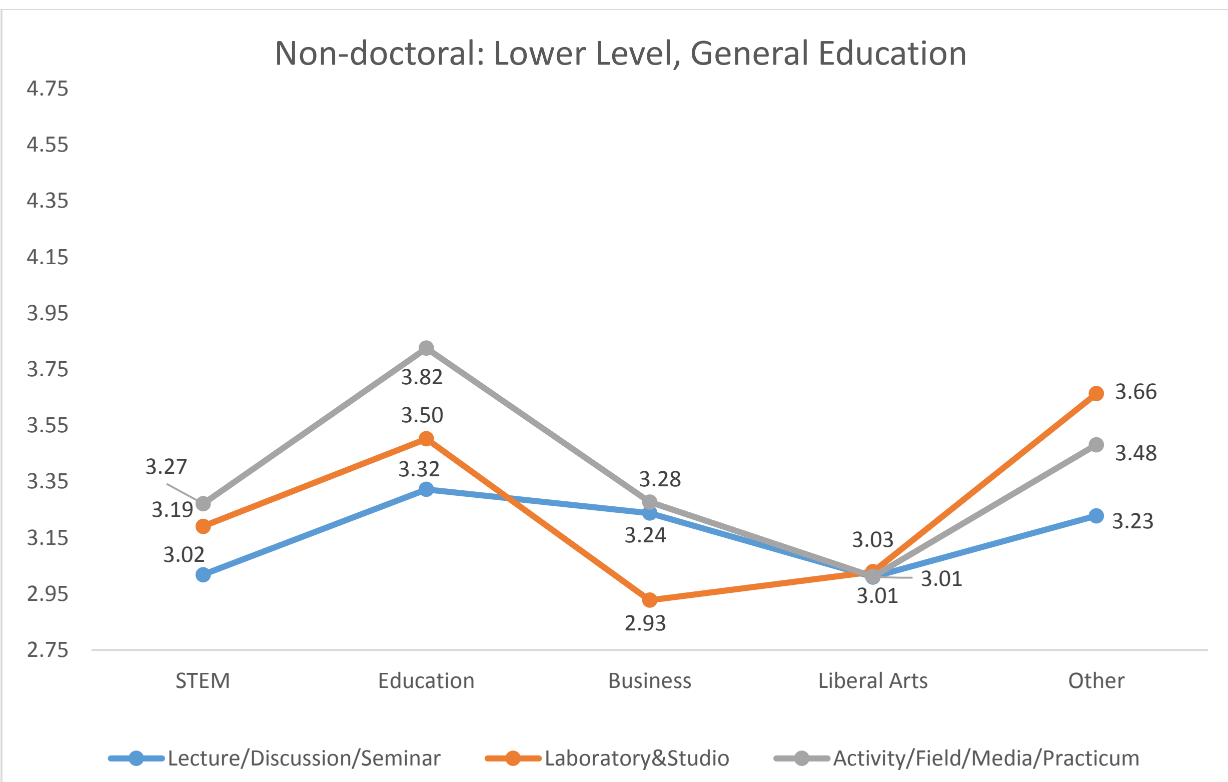


Figure 7 *Non-Doctoral: Upper Level General Education Student Motivation*



The mean ratings of motivation for STEM, Education, Business, and Liberal Arts determined by combining course type and student type were 3.35 (SE=.02), 3.53 (SE=.02), 3.31 (SE=.02), and 3.26 (SE=.05) respectively. Using the Scheffé statistic, students in Education disciplines had higher levels of motivation than all other disciplines, while Liberal Arts students had the lower levels of motivation than all other disciplines. The mean ratings for Lecture/Discussion/Seminar, Laboratory & Studio, and Activity/Field/Media/Practicum determined by combining discipline and student type were 3.34 (SE=.01), 3.46 (SE=.03), and 3.43 (SE=.02) respectively. Students had higher levels of motivation in Laboratory & Studio course types and lower levels of motivations in Lecture/Discussion/Seminar course types. The mean ratings for lower level general education, lower level specialization, upper level general education, upper level specialization, and graduate determined by combining course type and discipline were 3.17 (SE=.01), 3.49 (SE=.02), 3.27 (SE=.03), 3.54 (SE=.02), and 3.58 (SE=.04) respectively. In general, lower level general education course students had the lowest levels of student motivation while students in specialization or graduate level courses were more motivated than their general education counterparts regardless of level. Surprisingly, the mean difference between lower level specialization and graduate/professional was not significant.

Research Question 4

A three-way analysis of variance (ANOVA) was conducted to evaluate the relationship between overall student motivation and course type, discipline, and student type at doctoral granting institutions. The means and standard deviations for student motivation for course type, discipline and student type are presented in Appendix C. The results of the three-way ANOVA are reported in Table 4.9.

Table 4.9 *Three-Way Analysis of Variance Output for Student Motivation at Doctoral Institutions*

Dependent Variable: I really wanted to take this course regardless of who taught it (Doctoral)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4698.488a	72	65.257	274.533	.000
Intercept	15874.333	1	15874.333	66782.659	0.000
Discipline	240.116	4	60.029	252.540	.000
Course Type	8.421	2	4.210	17.713	.000
Student Type	52.471	4	13.118	55.186	.000
Discipline*Course Type	147.565	8	18.446	77.600	.000
Discipline*Student Type	69.812	16	4.363	18.356	.000
Course Type*Student Type	5.174	8	.647	2.721	.005
Discipline*Course Type*Student Type	73.066	30	2.436	10.246	.000
Error	32522.546	136821	.238		
Total	1607106.078	136894			
Corrected Total	37221.034	136893			

a. R Squared = .126 (Adjusted R Squared = .126)

Note: No class mean score for motivation were provided by the IDEA Center for 136 classes. Therefore, total sample size for the model is 136,894 compared to 137,030 displayed in Table 4.5.

The highest order interaction term of discipline, course type, and student type on student motivation was significant (*F* (30, 136,821) = 2.436, *p* = <.001, η^2 = .002). Although the highest order interaction term was significant, it is important, on average, to report the main effects. Students reported significantly different levels of student motivation by discipline (*F* (4, 136,821) = 252.540, *p* = <.001, η^2 = .007) at doctoral institutions. In addition, students reported significantly different levels of motivation by course type (*F* (2, 136,821) = 17.713, *p* = <.001, η^2 = .000). Finally, a significant difference for student motivation amongst student type (*F* (4, 136,821) = 55.186, *p* = <.001, η^2 = .002) existed. The highest order interaction term of discipline, course type, and student type on student motivation was significant (*F* (30, 136,821) = 2.436, *p* = <.001, η^2 = .002). When examining effect size (η^2) for each of the independent variables (discipline, course type, and student type), as well as all four interaction terms, and using the cut-off values provided in Cohen (1992) as a guide to determine the strength of the effect size, no notable effect size was observed ($r < .10$) from the factors in the model. In other words, no one independent variable, or interaction term, accounted for more than 1% of the total variance among student motivation.

The three-way interaction term of course type, discipline, and student type was found to be statistically significantly different. This finding indicated that the pattern of differences, noted above, between course type, discipline, and student type was significantly altered when all three factors were considered on doctoral granting institutions. After examining the various mean score plots, the interaction effect was present across all levels of the independent variables. While motivation seemed highest at the Graduate/Professional student type, and Lecture/Discussion/Seminar motivation levels seem lower, and the severity of the difference between the other levels of independent variables was inconsistent. For example, Education, STEM, and Business motivation mean values were very close to each other across levels of student type; when looking at motivation mean values across course type, the differences are large. Figures 8 and 9 provide an illustration of the interaction effect between course type and discipline when holding student type constant.

Figure 8 *Doctoral: Lower Level General Education Student Motivation*

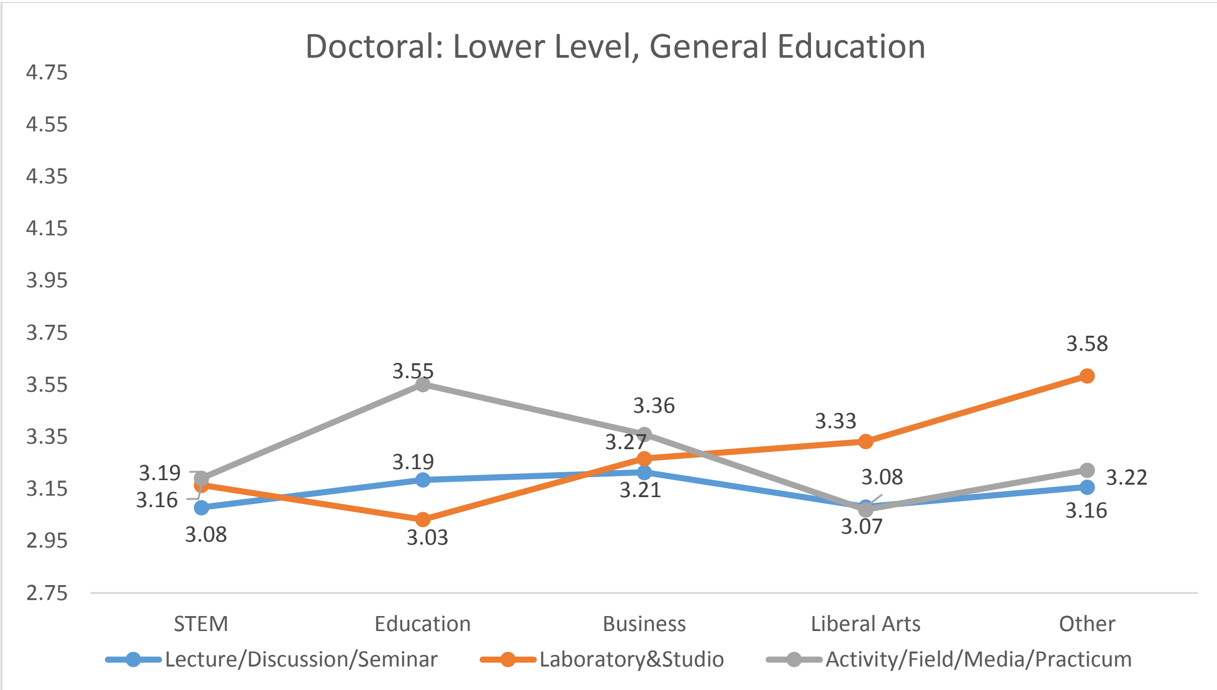
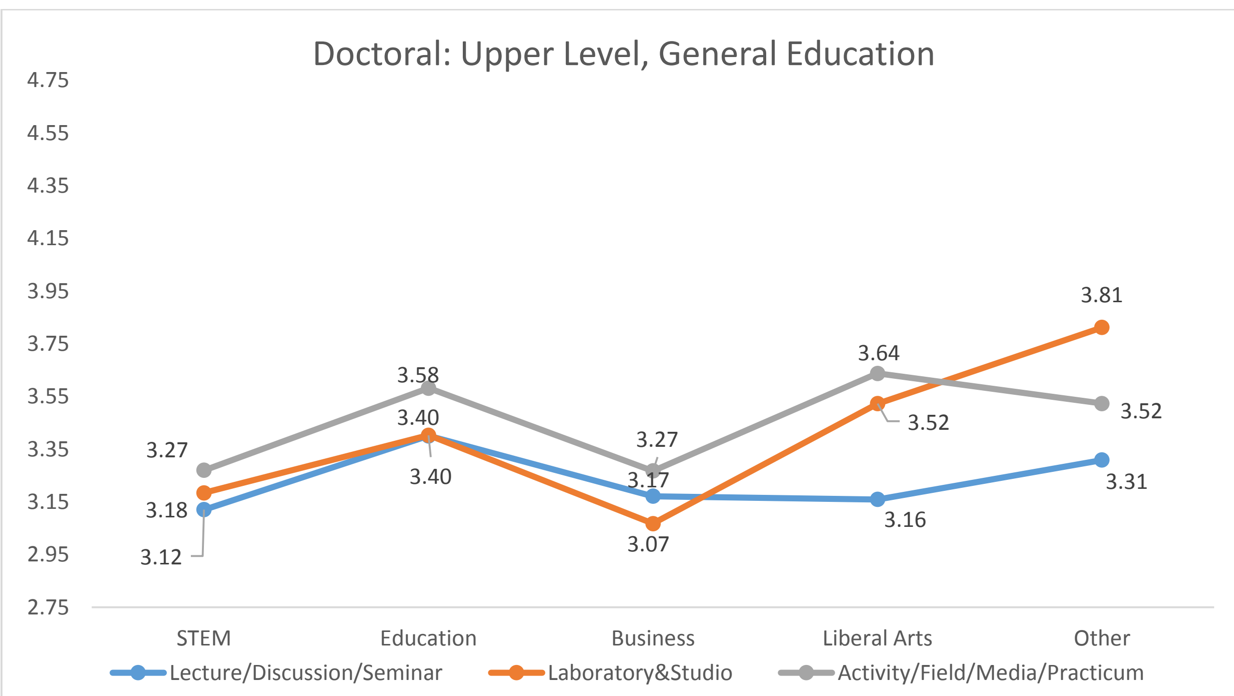


Figure 9 *Doctoral: Upper Level General Education Student Motivation*



The mean ratings of motivation for STEM, Education, Business, and Liberal Arts determined by combining course type and student type were 3.35 (SE=.01), 3.47 (SE=.01), 3.31 (SE=.02), and 3.45 (SE=.05) respectively. Using the Scheffé statistic, students in Education courses had higher levels of motivation than all other disciplines. No difference between STEM and Business existed. Liberal Arts students reported the lowest level of motivation. The mean ratings for Lecture/Discussion/Seminar, Laboratory & Studio, and Activity/Field/Media/Practicum determined by combining discipline and student type were 3.36 (SE=.01), 3.47 (SE=.02), and 3.48 (SE=.02) respectively. Students had higher levels of motivation in both Laboratory & Studio and Activity/Field/Media/Practicum course types compared to Lecture/Discussion/Seminar students. Lecture/Discussion/Seminar students had the lowest levels of motivation while Laboratory & Studio students reported the highest levels of motivation. The mean ratings for lower level general education, lower level specialization, upper level general education, upper level specialization, and graduate determined by combining course type and discipline were 3.22 (SE=.02), 3.48 (SE=.02), 3.36 (SE=.02), 3.55 (SE=.02), and 3.57 (SE=.04) respectively. Lower level general education course students had the lowest levels of student motivation while students in specialization or graduate level courses were more motivated than their general education counterparts regardless of level. No difference between lower level specialization and upper level specialization exists.

Research Question 5

A partial correlation between overall rating of teaching effectiveness and student motivation was conducted with the student type treated as a co-variate. A partial correlation test produced a correlation coefficient equal to .297. Partial correlations by institutional type resulted in similar values, .298 and .296 for doctoral and non-doctoral institutions respectively. An overall correlation coefficient of .297 equates to a r^2 value of .09. In summary, across all the class means, less than ten percent of the variance on motivation and teaching effectiveness was in common while over ninety percent was not common. A very weak relationship between motivation and student ratings of instruction existed.

Summary

This chapter presented the results of the statistical analyses that were conducted to address the research questions and hypotheses. The major findings of the study are as follows:

1. Students reported significant differences of student ratings of instruction by course type and discipline at non-doctoral granting institutions. The findings were similar at doctoral granting institutions, although student type was found to be significantly significant for doctoral granting institutions. That being said, the models explained very little of the variability between student ratings of instruction (r^2 values ranged between .015 and .023). Not surprisingly, effect sizes were negligible across all significant findings.
2. Students reported significant differences of motivation by course type, discipline, and student type at both non-doctoral and doctoral granting institutions. Again, the models explained little of the variability between student levels of motivation (r^2 values ranged between .126 and .166). Again, effect sizes were negligible across all significant findings.
3. Lastly, a very weak relationship between student motivation and student ratings of instruction exists.

Chapter 5 - DISCUSSSION

Purpose of the Study

The purpose of this study was to examine institution type, course type, discipline, and student type to determine how these variables affect student ratings of instruction both individually and collectively. In addition, the connection between student ratings of instruction and motivation was explored. The findings from this study provided insight into the discrepancies between student ratings among differences in course type, discipline, and student type. Likewise, the results from chapter four demonstrated the lack of connection between student ratings of instruction and motivation.

Summary of the Results

The results of this study can be summarized as follows. First, at both doctoral and non-doctoral level institutions, course type (lecture, lab, other) and discipline affected overall student ratings of instruction. At doctoral institutions, student type (lower level general education, upper level major, graduate, etc.) affected student ratings of instruction. Secondly, all three factors (course type, student type, and discipline) had an effect on student motivation at both institution types. These findings are statistically significant; however, they have minimal practical significance, as noted by the effect size. Lastly, there was a weak relationship between the level of student motivation and student ratings of instruction. A full discussion of the implications of these research findings are presented below. They are discussed within the context of teaching and research.

Discussion of Research Questions

Research Question 1 Discussion:

"When looking at non-doctoral Degree-Granting institutions, is there an effect of course type, discipline, and course level on overall ratings of instruction?"

The results demonstrated that significant differences existed by discipline and by course type. No differences were detected by student type. In addition, the interaction between discipline, course type, and student affected student ratings of instruction. Each of these findings, as well as implications for practice, is included in the following paragraphs.

Students in science, technology, engineering, and mathematics (STEM) disciplines received the lowest class mean scores while education related disciplines received the highest class mean scores. Business, liberal arts, and disciplines considered as "other" showed mixed results, i.e. some were rated higher and some were rated lower. This finding is somewhat consistent with Centra's (2003; 2009) findings of STEM disciplines being rated lower while social sciences were rated higher. Education was the closest social science related discipline explored in this study.

Course type yielded differences between the three course types used in this study: (1) lecture/discussion/seminar, (2) laboratory and studio, and (3) activity/field/media/practicum. The lecture/discussion/seminar showed the lowest mean scores while laboratory and studio course types showed the highest mean scores even when holding discipline and student type constant. It was not surprising to the researcher that class types requiring a significant amount of hands-on application, common in a laboratory or studio course, would receive higher marks related to teaching effectiveness compared to course types with a more theoretical framework.

The effect of student type (lower-level general education versus graduate for example) showed no difference. In other words, at non-doctoral-degree granting institutions, teaching effectiveness was not affected by whether an instructor taught a first-year survey course or if it was an upper-level capstone experience for major students. The stability of student ratings, regardless of student type is consistent with the work of Marsh and Overall (1981) and Scherr and Scherr (1990).

Lastly, the interaction between course type, discipline, and student type revealed that teaching effectiveness was influenced by these three factors. In other words, teaching effectiveness cannot be assessed independently from other factors. As discussed earlier, activity/field/media/practicum showed the least consistency across factors compared to other course types. What is consistent regarding activity/field/media/practicum course types is that they are consistently rated the lowest at the graduate level, regardless of discipline. With the theoretical emphasis at the graduate level, connecting the theoretical base with practical applications may be difficult for most instructors. When looking at the interaction effect across student types, the interaction effect is relatively constant, i.e. same pattern holds across student types. This result adds support to the findings of no difference in student ratings of instruction by student type.

The implications for practice are limited, but noteworthy. For institutions that primarily serve baccalaureate students, teaching strategies designed for survey courses should have utility for upper-level and graduate-level courses. The effect of the interaction between all three factors (student type, course level, and discipline) cannot be ignored. College and university administrators need to be careful before implementing "one-size fits all" instructional strategies. A one-size fits all approach, where instructional strategies are consisted regardless of level may work at one institution, but the generalizability across similar types of non-doctoral-granting institutions is limited. Instead institutions should devote resources to develop instructional strategies that work best at their specific institutions.

Research Question 2 Discussion:

"When looking at Doctoral Degree-Granting institutions, is there an effect of course type, discipline, and course level on overall ratings of instruction"

The results demonstrated that significant differences existed by discipline, course type, and student type. In addition, the interaction between discipline, course type, and student affected student ratings of instruction. Each of these findings, as well as implications for practice, is included in the following paragraphs.

Students in science, technology, engineering, and mathematics (STEM) disciplines received the lowest class mean scores. Education, business, liberal arts, and disciplines considered as "other" showed mixed results, i.e. some were rated higher and some were rated lower. When compared to non-doctoral-granting institutions, STEM courses were again rated the lowest but education courses were no different than any other discipline.

Course type yielded differences between the three course types used in this study. As expected, lecture/discussion/seminar showed the lowest mean scores, significantly lower than the other two course types. Activity/Field/Media/Practicum course types were significantly higher than the other two course types. These findings are slightly different from the non-doctoral findings, but regardless of institutional type, course types with more applied learning applications received higher scores related to teaching effectiveness. These findings are in contrast to Aleamoni (1999) who concluded that differences by course type were unsupported in the research.

The effect of student type (lower-level general education versus graduate for example) showed significant differences. In general, as students progressed through their coursework, they reported higher mean scores related to teaching effectiveness with students in upper-level specialization and graduate courses reporting the highest mean scores. This finding is consistent with the finding s of Santhanam and Hicks (2002) who reported higher mean scores for teaching effectiveness for upper-division courses compared to lower-division courses

Lastly the interaction between course type, discipline, and student type revealed that teaching effectiveness was influenced by these three factors. In other words, teaching effectiveness cannot be assessed independently from other factors. Contrary to the findings at the non-doctoral level, the interaction effect is more sporadic, i.e. no general pattern was identified. The significant difference by student type could account for this difference when compared to non-doctoral-granting institutions.

When considering implications for practice, compared to the findings from non-doctoral granting institutions, student type was a significant finding. Because of this finding, teaching practices for baccalaureate serving institutions may or may not be transferable to doctoral granting institutions. However, not unlike non-doctoral serving institutions, the interaction term between course type, student type, and discipline is significant. Thus regarding strategies to increase teaching effectiveness, the transferability to a larger breadth of institutions seems plausible. A point of caution should be given considering the large amount of variability unaccounted for in the model used for this study. In other words, almost 98 percent of the differences between student ratings of instruction at doctoral-granting institutions are unaccounted for after adjusting for course type, disciple, and student type. Within institutional variance, as well as instructor type, might account for more of the variance between student ratings of instruction at doctoral-granting institutions.

Research Question 3 Discussion:

"When looking at non-doctoral Degree-Granting institutions, is there an effect of course type, discipline, and course level on students' motivation to take the class?"

The results demonstrated that significant differences existed by discipline, course type, and student type. In addition, the interaction between discipline, course type, and student affected student motivation. Each of these findings, as well as implications for practice, is discussed below.

Students in Liberal Arts disciplines received the lowest class mean scores while education related disciplines received the highest class mean scores. Business, STEM, and disciplines considered as "other" showed mixed results.

Course type yielded differences between the three course types used in this study. Not surprisingly lecture/discussion/seminar showed the lowest mean scores, while laboratory and studio course types showed the highest mean score. It is not surprising that course types with more applied modality received high marks related to student motivation considering the level of student involvement in laboratory and studio courses. Using Bandura's (1977) self-efficacy model, the generality of laboratory or studio courses would be more transferable than a more abstract theoretical concept from a lecture/discussion/seminar course type.

The effect of student type (lower-level general education versus graduate for example) showed significant differences. Overall, students in specialization or graduate courses had higher levels of motivation compared to their general education counterparts. This finding seems obvious, yet the fact that lower level specialization classes received higher mean scores is perplexing. Only graduate students showed higher levels of motivation than lower-level specialization students. From a self-efficacy standpoint, this fact is puzzling. According to Bandura (1977) higher levels of self-efficacy lead to higher levels of motivation. Therefore, using a self-efficacy framework regarding student motivation, one conclusion could be that upper-level academic proficiencies do not provide the meaningful experiences needed to build self-efficacy among undergraduate students at non-doctoral-granting institutions.

Lastly, the interaction between course type, discipline, and student type revealed that student motivation was influenced by these three factors. Student motivation cannot be assessed independently from other factors. As discussed earlier, in general, motivation was higher in specialization and graduate level courses when holding other factors constant. Considering the importance of mastery experiences in building self-efficacy (Bandura, 1977; 1986; 2002), classes where students had positive experiences to draw upon prior to enrolling, higher motivational levels would be expected compared to general education where students may have little or no positive experiences to draw upon. After adding discipline and course type to the equation, motivation levels varied considerably. With little change in the motivational level of STEM disciplines across student types, one could consider that the rigorous coursework in STEM disciplines might lead students to believe they are just as capable (if not more) than their peers in non-specialization courses, and therefore they remain motivated regardless of course type, i.e. strong self-efficacy via vicarious experiences.

Bandura’s self-efficacy model is important when considering implications for practice. According to Bandura (1977; 1986), mastery experiences are the most important source of information for making self-efficacy judgments. Therefore, as illustrated in this research, as student progress from lower-level to upper-level mean scores related to motivation increase and more motivated students perform better academically (Chemers et al, 2001). In addition, being cognizant of motivational differences amongst students is important from an institutional perspective because institutions could develop training and professional development opportunities for instructors of lower motivational areas, specifically in liberal arts disciplines or lower-level survey courses.

Research Question 4 Discussion:

"When looking at Doctoral Degree-Granting institutions, is there an effect of course type, discipline, and course level on students' motivation to take the class?"

The results demonstrated that significant differences existed by discipline, course type, and student type. In addition, the interaction between discipline, course type, and student affected student motivation. Each of these findings, as well as implications for practice, is included in the following paragraphs.

Students in liberal arts disciplines received the lowest class mean scores while education related disciplines received the highest class mean scores. Business and STEM disciplines had equal levels of student motivation. These findings were consistent with the non-doctoral degree findings, an important similarity for college and university administrators.

Course type yielded differences between the three course types used in this study. Lecture/discussion/seminar showed the lowest mean scores while laboratory and studio course types showed the highest mean scores holding discipline and student type constant. One would expect that course types with more applied delivery would receive higher marks related to student motivation considering the level of student involvement in laboratory and studio courses is substantially higher than a lecture or discussion course. Again these findings are in line with non-doctoral degree findings.

The effect of student type (lower-level general education versus graduate for example) showed significant differences. Generally, students in specialization or graduate courses had higher levels of motivation compared to their general education counterparts. Compared to non-doctoral-granting institutions, the level (upper or lower) did not show any differences in student motivation.

Lastly the interaction between course type, discipline, and student type revealed that student motivation was influenced by these three factors. Because of this interaction effect, discussing motivational factors independent of the other two is misleading. Similar to the findings at the non-doctoral level, motivation seemed to increase as students had more positive experiences to draw upon in their academic career, i.e. mastery experiences. Considering the academic rigor in STEM graduate fields and Schunk's (1984) findings regarding preparedness and the effect they have on self-efficacy and ultimately motivation, it is not overly surprising that STEM fields showed the highest levels of motivation compared to all disciplines at the graduate level.

When considering implications for practice, not unlike non-doctoral granting institutions, the interaction term between course type, student type, and discipline is significant. Thus regarding strategies to increase student motivation, the transferability to a larger breadth of institutions seems plausible. A point of caution should be given considering the large amount of variability unaccounted for in the model used for this study.

Research Question 5 Discussion:

"Is there a relationship between students' overall ratings of instruction and their motivation to take the class when course level is held constant?"

A very weak correlation between student ratings of instruction and student motivation was detected. Less than ten percent of the variance among student ratings of instruction and student motivation was common, with over ninety percent unique. Surprisingly, this finding is good for those that place considerable weight in student ratings of instruction (Centra, 2003; 2009; Marsh, 1984) and their validity. If a stronger relationship existed, faculty and administrators that are critical of using student ratings of instruction to gauge teaching effectiveness could argue that using student ratings of instruction are meaningless for measuring teaching effectiveness. As opposed to developing instructional strategies, colleges and universities would be better suited to devote additional resources to mentoring and coaching skills for instructors, i.e. building self-efficacy among students, the key to student motivation according to Bandura.

Considerations for Future Research

Based on the results in Chapter 4, and the additional explanation provided in this chapter, several important factors are yet unaccounted for in literature. The following recommendations for future research include:

1. A qualitative study profiling the self-efficacy of students in higher education across a variety of student types, i.e. students in general education lower level courses versus students in graduate level course. Ideally, profiles by institution type (non-doctoral-granting versus doctoral-granting) would be ideal.

2. A quantitative study using primary data, i.e. researcher gathered data through survey/testing and/or observation, examining the relationship between student motivations and teaching effectiveness should be conducted and compared to the secondary data study conducted in this research. Included in the comparison would be type of instructor (graduate teaching assistant versus tenured/tenure-track versus non-tenured for example) and what effect, if any, this has on student ratings of instruction, a limitation of the current study.
3. A supplementary study should be undertaken to examine additional factors related to differences in teaching effectiveness and student motivation since the current study accounted for so little of the variance associated with student ratings of instruction and student motivation. Other factors might include additional student demographics including age of students and even gender. In addition, as illustrated above, controlling for instructor type would be beneficial.
4. While the current study was primarily concerned with teaching effectiveness and student motivation, a better understanding of student characteristics and how they influence teaching effectiveness is needed. A study that accounts for academic achievement, even on a smaller scale to the current study, is needed.

Overall Recommendations for Practice

After thoughtful reflection three main takeaways for practice were identified by the researcher. First, instructors should building self-efficacy building skills into the curriculum early in the academic term. Giving students, regardless of discipline and course type, the opportunity to build mastery experiences in the specific subject matter is important, especially in courses or disciplines where students have had difficulty in the past. Since self-efficacy is the key to improving motivation of struggling students (Margolis & McCabe, 2003) and students with high self-efficacy beliefs are most likely to persist in the face of difficult material as opposed to those with low self-efficacy beliefs (Bandura, 1982), building up the self-efficacy levels of students is critical. For example, with a class of primarily first-year students, as opposed to writing a large research paper or completing a complex group project towards the end of the semester, the project could be broken out over a number of assignments. For instance the first assignment could be to identify primary or secondary data sources to be used in the study, familiarizing them with the library. The second assignment could be producing an annotated bibliography. The third assignment would be producing a detailed outline. This provides the opportunity for students to complete smaller tasks, receive timely feedback, and build a solid base of success to draw upon when faced with more difficult tasks later in the process.

The second implication for practice is for instructors to better utilize classroom demographics. As discovered in this study, student ratings of instruction and motivation were higher for students in major related courses at both lower-level and upper-level sections. Especially when instructing lower-level courses, instructors need to understand why individuals are enrolled in the specific course. Is it for general education or is it part of their major/field of study? If there is a clear majority one way or another, instructors need to incorporate this knowledge in their lesson planning and delivery method. Conversely, if the course is split with roughly equal proportions taking it for general education as those taking it for major/field of study, the instructor needs to use a hybrid approach to be as inclusive as possible to the breadth of students in his or her classroom. Thankfully, technology has allowed instructors the ability to gather this information early in the academic term via the student information systems commonly in place at institutions. If no centralized system is present, or more information is needed, a simple straw poll or general information questionnaire can be developed and administered early in the academic term.

The last recommendation for practice, and arguably the one that might have the largest push-back from faculty, especially tenure-track or tenured faculty, is the need for colleges and universities to provide instruction on pedagogy. Included in that training, especially for college and university faculty who have been instructing students for a long period of time, is the need for professional development on best practices and how the modern student receives and processes information in the modern learning environment. Considering the warning issued by Feldman (2007) as to the increased priority colleges and universities are placing on developing, identifying, and recognizing good instruction, it would be fitting for colleges and universities to devote adequate resources in "teaching how to teach". Workshops led by highly respected instructors both in the eyes of faculty as well as students, national experts on pedagogy, and additional resources to institutional teaching and learning centers are a few of things that can be used to increase the institutional commitment to improving instruction. When adding differences by institutional type, course type, and student type, it is not enough for institutions to say that a faculty member learned how to teach during his/her graduate school experience, especially since the institution could be drastically different from the institution he/she is currently employed. Institutions need to understand how these factors affect learning and how they can help students meet their educational goals by improving instruction.

Summary

The results from this study found that factors like institutional type, course type, discipline, and student type have an effect on both student motivation as well as student ratings of instruction. However, the relationship between student ratings of instruction and student motivation is limited at best. Therefore, institutions need to devote adequate resources of time and money to increasing both instructional quality and self-efficacy among their respective student bodies. Despite these general takeaways, a more thorough understanding of differences in both student ratings of instruction and motivation is needed.

References

Agresti, A. & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924-1988. *Journal of Personnel Evaluation in Education*, *13*(2), 153-166.

Arias, J. J. & Walker, D. M. (2004). Additional evidence on the relationship between class size and student performance. *Journal of Economic Education, 35*(4), 311-329.

Bain, K. (2004). *What the best college teachers do.* Cambridge, MA: Harvard University Press.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review, 84*(2), 191-215.

Bandura, A. (1978). The self-system in reciprocal determinism. *American Psychologist, 33*(4), 344-358.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122-147.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 44*(9), 1175-1184.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, *28*(2), 117-148.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*, 1-26.

Bandura, A. (2002). Social cognitive theory in cultural context. *Applied Psychology: An International Review, 51*(2), 269-290.

Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-efficacy beliefs of adolescents* (pp. 307-337) Information Age Publishing.

Bandura, A. & Wood, R. (1989). Effect of perceived controllability and performance standards on self-regulation of complex decision making. *Journal of Personality and Social Psychology, 56*(5), 805-814.

Basow, S.A. & Montgomery, S. (2005). Student ratings and professor self-ratings of college teaching: Effect of gender and divisional affiliation. *Journal of Personnel Evaluation in Education, 18*, 91-106.

Bausell, R. B., Schwartz, S., & Purohit, A. (1975). An examination of the conditions under which various student rating parameters replicate across time. *Journal of Educational Measurement, 12*(4), 273-280.

Benton, S. L. & Cashin, W. E. (2012). *Student ratings of teaching: A summary of research and literature.* (No. 50). Manhattan, KS: IDEA Center.

Bettinger, E. P. & Long, B. T. (2010). Does cheaper mean better? The impact of using adjunct instructors on student outcomes. *The Review of Economics and Statistics, 92*(3), 598-613.

Braskamp, L.A. & Ory, J.C. (1994). *Assessing faculty work* (1st ed.). San Francisco, CA: Jossey-Bass Inc.

Braskamp, L.A., Brandenburg, D. C., & Ory, J. C. (1984). *Evaluating teaching effectiveness: A practical guide*. Beverly Hills, CA: Sage Publications.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

Campbell, D.T. & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin Company.

Cashin, W. E. (1990). Students do rate different academic fields differently. *New Directions for Teaching and Learning, 43*, 113-121.

Cashin, W. E., Noma, A., & Hanna, G. S. (1987). *IDEA technical report no. 6. comparative data by academic field*. Manhattan, KS: Center for Faculty Evaluation and Development.

Cashin, W. E. & Slawson, H. M. (1977). *Description of data base.* (No. 3). Manhattan, KS: IDEA Center.

Centra, J. A. (1978). Using student assessments to improve performance and vitality. *New Directions for Institutional Research, 20*, 31-49.

Centra, J. A. (1993). *Reflective faculty evaluation. Enhancing teaching and determining faculty effectiveness*. San Francisco, CA: Jossey-Bass Publishers.

Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education, 44*(5), pp. 495-518.

Centra, J. A. (2009). *Differences in responses to the student instructional report: Is it bias?*. Princeton, NJ: Educational Testing Service.

Centra, J. A., & Creech, F. R. (1976). *The relationship between student, teacher, and course characteristics and student ratings of teacher effectiveness.* Princeton, NJ: Educational Testing Service.

Centra, J.A. & Gaubatz, N.B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education, 71*(1), 17-33.

Chemers, M. M., Hu, L., & Garcia, B. F. (2001). Academic self-efficacy and first-year college student performance and adjustment. *Journal of Educational Psychology, 93*(1), 55-64.

Cheng, J. H. S. & Marsh, H. W. (2010). National student survey: Are differences between universities and courses reliable and meaningful? *Oxford Review of Education, 36*(6), 693-712.

Clayson, D. E. (2008). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education, 31*(1), 16-30.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Costin, F., Greenough, W. T., & Menges, R. J. (1971). Student ratings of college teaching: Reliability, validity, and usefulness. *Review of Educational Research, 41*(5), 511-535.

Culver, S. (2010). Course grades, quality of student engagement, and students' evaluation of instructor. *International Journal of Teaching and Learning in Higher Education, 22*(3), 331-336.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12*(2), 163-170.

Department of Homeland Security. (2014, October 5). STEM designated degree programs. Retrieved from http://www.ice.gov/sevis/stemlist.htm

d'Apollonia, S. & Abrami, P.C. (1997). Navigating student ratings of instruction. *American Psychologist, 52*(11), 1198-1208.

Eagen Jr., M. K. & Jaeger, A. J. (2008). Closing the gate: Part-time faculty instruction in gatekeeper courses and first-year persistence. *New Directions for Teaching and Learning, 115*, 39-53. doi:10.1002/tl.324

Educational Testing Service. (2010a). *SIR II student instructional report II: Comparative data community colleges, two-year colleges and technical institutions*. Educational Testing Service.

Educational Testing Service. (2010b). *SIR II student instructional report II: Comparative data four-year colleges and universities.* Educational Testing Service.

El Hassan, K. (2009). Investigating substantive and consequential validity of student ratings. *Higher Education Research and Development, 28*(3), 319-333.

Feldman, K. A. (1976). The superior college teacher from the students' view. *Research in Higher Education, 4*(4), 69-111.

Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers: What we know and what we don't. *Research in Higher Education, 9*(3), 199-242.

Feldman, K. A. (1984). Class size and college student's evaluation of teachers and courses: A closer look. *Research in Higher Education, 21*(1), 45-116.

Feldman, K. A. (1987). Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness: A review and exploration. *Research in Higher Education, 26*(3), 227-298.

Feldman, K. A. (1992). College students' views of male and female college teachers: Part I: Evidence from the social laboratory and experiments. *Research in Higher Education, 33*(3), 317-375.

Feldman, K. A. (1993). College students' views of male and female college teachers: Part II: Evidence from students' evaluations of their classroom teachers. *Research in Higher Education, 34*(2), 151-211.

Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. In R. P. Perry, & J. C. Smart (Eds.). *The scholarship of teaching and learning in higher education: An evidence based perspective.* Dordrecht, The Netherlands: Springer.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Los Angeles: Sage Publications.

Freeman, H. R. (1994). Student evaluations of college instructors: Effects of type of course taught, instructor gender and gender role, and student gender. *Journal of Educational Psychology, 86*, 627-630. doi:10.1037/0022-0663.86.4.627

Gillmore, G. M., Kane, M. T., & Naccarato, R. W. (1978). The generalizability of student ratings of instruction: Estimation of the teacher and course components. *Journal of Educational Measurement, 15*(1), 1-13.

Gilmore, D. C., Swerdlik, M. E., & Beehr, T. E. (1980). Effects of class size and college major on student ratings of psychology courses. *Teaching of Psychology, 7*(4), 210-214.

Goldberg, G. & Callahan, J. (1991). Objectivity of student evaluations of instructors. *Journal of Education for Business, 66*(6), 377-378.

Goldman, L. (1993). On the erosion of education and the eroding foundations of teacher education (or why we should not take student evaluation of faculty seriously). *Teacher Education Quarterly, 20*(2), 57-64.

Graham, S. & Weiner, B. (1996). Theories and principles of motivation. In D. C. Berliner, & R. C. Calfee (Eds.), *Handbook of educational psychology* (6th ed., pp. 63-84) Routledge.

Greenwald, A. G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist, 52*(11), 1182-1186.

Greenwald, A.G. & Gillmore, G.M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist, 52*(11), 1209-1217.

Guder, F., Malliaris, M., & Jalilvand, A. (2009). Changing the culture of school: The effect of larger class size on instructor and student performance. *American Journal of Business Education, 2*(9), 83-90.

Hellman, C. M. (1998). Faculty evaluation by students: A comparison between full-time and adjunct faculty. *Journal of Applied Research in the Community College, 6*, 45-50.

Hoffman, F. & Oreopoulos, P. (2006). Professor qualities and student achievement. *In Review,*

Holton, E. F. I., Bates, R. A., Bookter, A. I., & Yamkovenko, V. B. (2007). Convergent and divergent validity of the learning transfer system inventory. *Human Resource Development Quarterly, 18*(3), 385-419. doi:10.1002/hrdq.1210

Hoyt, D. P., & Lee, E. J. (2002a). *Basic data for the revised IDEA system.* (No. 12). Manhattan, KS: IDEA Center.

Hoyt, D. P., & Lee, E. J. (2002b). *Disciplinary differences in student ratings.* (No. 13). Manhattan, KS: IDEA Center.

Huemer, M. W. (2005). Student evaluations: A critical review. Retrieved from http://home.sprynet.com/~owl1/sef.htm#_14

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38*(4), 319-342.

Kember, D. & Leung, D. Y. P. (2011). Disciplinary differences in student ratings of teaching quality. *Research in Higher Education, 52*, 278-299. doi:10.1007/s11162-010-9194-z

Kohlberg, L. (1973). The claim to moral adequacy of a highest stage of moral judgment. *Journal of Philosophy, 70*(18), 630-646.

Kuehl, R. O. (2000). *Design of experiments: Statistical principles of research design and analysis* (2nd ed). Pacific Grove, CA: Brooks/Cole.

Kulik, J. A. & Kulik, C. C. (1974). Student ratings of instruction. *Teaching of Psychology, 1*(2), 51-57.

Landrum, R. E. (2009). Are there instructional differences between full-time and part-time faculty? *College Teaching, 57*(1), 23-26.

Latham, G. P. & Yukl, G. A. (1975). A review of research on the application of goal setting in organizations. *The Academy of Management Journal, 18*(4), 824-845.

Linnenbrink, E. A. & Pintrich, P. R. (2002). Motivation as an enabler for academic success. *School Psychology Review, 31*(3), 313-327.

Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance, 3*, 157-189.

Locke, E. A. (1975). Personnel attitudes and motivation. *Annual Review of Psychology, 26*, 457-480.

Locke, E. A. & Latham, G., P. (1990). Work motivation and satisfaction: Light at the end of the tunnel. *Psychological Science, 1*(4), 240-246.

Locke, E. A. & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation. *American Psychologist, 57*(9), 705-717.

Locke, E. A., Latham, G. P., & Erez, M. (1988). The determinants of goal commitment. *The Academy of Management Review, 13*(1), 23-29.

Lunenburg, F. C. (2011). Self-efficacy in the workplace: Implications for motivation and performance. *International Journal of Management, Business, and Administration, 14*(1), 1-6.

Margolis, H. & McCabe, P. P. (2003). Self-efficacy: A key to improving the motivation of struggling learners. *Preventing School Failure, 47*(4), 162-169.

Marsh, H. W. (1982). Validity of students' evaluations of college teaching: A multitrait-multimethod analysis. *Journal of Educational Psychology, 74*(2), 264-279.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology, 76*(5), 707-754.

Marsh, H. W. (1987). Students' evaluation of university teaching: Research findings, methodological issues and directions for future research. *International Journal of Educational Research, 11*(3), 253-388.

Marsh, H. W. (1991). Multidimensional students' evaluations of teaching effectiveness: A test of alternative higher-order structures. *Journal of Educational Psychology, 83*(2), 285-296.

Marsh, H. W. (2001). Distinguishing between good (useful) and bad workloads on students' evaluations of teaching. *American Educational Research Journal*, *38*(1), 183-212.

Marsh, H. W. (2007). Do university teachers become more effective with experience? A multilevel growth model of students' evaluations of teaching over 13 years. *Journal of Educational Psychology, 99*(4), 775-790. doi:10.1037/0022-0663.99.4.775

Marsh, H. W., & Dunkin, M. (1992). Students evaluations of university teaching: A multidimensional perspective. In G. A. Marcoulides, & R. E. Schumacker (Eds.), *Higher education handbook on theory and research* (8th ed., pp. 143-224). New York: Agathon.

Marsh, H. W. & Hocevar, D. (1984). The factorial invariance of student evaluations of college teaching. *American Educational Research Journal, 21*(2), 341-366.

Marsh, H. W. & Overall, J. U. (1981). The relative influence of course level, course type, and instructor on students' evaluations of college teaching. *American Educational Research Journal, 18*(1), 103-112.

Marsh, H. W. & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American Psychologist, 52*(11), 1187-1197.

Marsh, H. W., & Roche, L. A. (2000). Effects of grading leniency and low workload on students' evaluations of teaching: Popular myths, bias, validity, or innocent bystanders? *Journal of Educational Psychology, 92*(1), 202-228.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50*(4), 370-396.

Maslow, A. H. (1954). *Motivation and personality.* New York: Harper.

McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist, 52*(11), 1218-1225.

Murray, H. G., Rushton, J. P., & Paunonen, S. V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology, 82*(2), 250-261.

Naftulin, D. H., Ware, J., J.E., & Donnelly, F. A. (1973). The doctor fox lecture: A paradigm of educational seduction. *Journal of Medical Education, 48*, 630-635.

Nelson-Laird, T.F., Niskode-Dossett, A.S. & Kuh, G.D. (2009). What general education courses contribute to essential learning outcomes. *The Journal of General Education, 58*(2), 65-84.

Nichols, M. L. & Day, V. E. (1982). A comparison of moral reasoning of groups and individuals on the "defining issues test". *The Academy of Management Journal, 25*(1), 201-208.

Oliver-Hoyo, M. (2008). Student evaluations that generate more questions than answers. *Journal of College Science Teaching, 38*(1), 37-39.

Ory, J.C. & Ryan, K. (2001). How do student ratings measure up to a new validity framework? *New Directions for Institutional Research, 109*, 27-44.

Overbaugh, R. C. (1998). The effect of course rigor on pre-service teachers' course and instructor evaluation. *Computers in the Schools, 14*(3-4), 13-23.

Peer, E. & Babad, E. (2012). The doctor fox research (1973) re-revisited: "educational seduction" ruled out. *Journal of Educational Psychology, In Review*

Piaget, J. (1964). Cognitive development in children: Piaget. *Journal of Research in Science Teaching, 2*, 176-`86.

Richardson, J. T. E. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment and Evaluation in Higher Education, 30*(4), 387-415.

Santhanam, E. & Hicks, O. (2002). Disciplinary, gender and course year influences on student perceptions of teaching: Explorations and implications. *Teaching in Higher Education, 7*(1), 17-31.

Santos, J. R. A. (1999). Cronbachs's alpha: A tool for assessing the reliability of scales. *Journal of Extension, 37*(2)

Scherr, F. C. & Scherr, S. S. (1990). Bias in student evaluations of teacher effectiveness. *Journal of Education for Business, 65*(8), 356-358.

Schmelkin, L. P., Spencer, K. J., & Gellman, E. S. (1997). Faculty perspective on course and teacher evaluations. *Research in Higher Education, 38*(5), 575-592.

Schunk, D. H. (1984). Self-efficacy perspective on achievement behavior. *Educational Psychologist, 19*, 48-58.

Seldin, P. (1993). The use and abuse of student ratings of professors. *Chronicle of Higher Education, 39*(46), A40.

Skinner, B. F. (1953). *Science and human behavior*. New York: The Free Press.

Spencer, K. J. & Schmelkin, L. P. (2002). Student perspective on teaching and its evaluation. *Assessment & Evaluation in Higher Education, 27*(5), 397-409.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99-103.

Theall, M. & Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? *New Directions for Institutional Research,* (109), 45-56.

Trout, P. (2000). Flunking the test: The dismal record of student evaluations. *Academe Online, 86*(4), 58-61.

Tschannen-Moran, M. & Hoy, A. W. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education, 23*(6), 944-956.

Umbach, P. D. (2007). How effective are they? Exploring the impact of contingent faculty on undergraduate education. *The Review of Higher Education, 30*(2), 91-123.

University of Delaware. (2013, April 11). The national study of instructional costs and productivity. Retrieved from http://www.udel.edu/IR/cost/

White, S. S. & Locke, E. A. (2000). Problems with the pygmalion effect and some proposed solutions. *Leadership Quarterly, 11*(3), 389-415.

Whitworth, J. E., Price, B. A., & Randall, C. H. (2002). Factors that affect college of business student opinion of teaching and learning. *Journal of Education for Business, 77*(5), 282-289. doi:10.1080/08832320209599677

Wollert, M. H., & West, R. F. (2000). Differences in student ratings of instructional effectiveness based on the demographic and academic characteristics of instructors. *Mid-South Educational Research Association,* Bowling Green, KY.

Wright, S. L. & Jenkins-Guarnieri, M. A. (2012). Student evaluations of teaching: Combining the meta-analyses and demonstrating further evidence for effective use. *Assessment & Evaluation in Higher Education, 37*(6), 683-699.

Zhao, J. & Gallant, D. J. (2012). Student evaluation of instruction in higher education: Exploring issues of validity and reliability. *Assessment & Evaluation in Higher Education, 37*(2), 227-235.

Zoller, U. (1992). Faculty teaching performance evaluation in higher science education: Issues and implications (A "cross-cultural" case study). *Science Education, 76*(6), 673-684.

Appendix A - IDEA Center Forms

IDEA Center Release Form

THE IDEA CENTER
T: 800.255.2757
T: 785.320.2400
F: 785.320.2424
301 South Fourth Street, Suite 200
Manhattan, KS 66502-6209
www.theideacenter.org

March 28, 2013

Insight.
Improvement.
Impact.®

Christopher Feit
Graduate Student
Department of Special Education, Counseling and Student Affairs
College of Education
1100 Mid-Campus Drive
Bluemont Hall
Kansas State University

Dear Mr. Feit:

This is in response to your request of March 28, 2013. You are given permission to include a copy of the IDEA Student Ratings Diagnostic Form and Faculty Information Form in the appendix of your dissertation. This permission is given under the following conditions:

1. that the surveys not be altered in any way;

2. that The IDEA Center be indicated as the surveys' source; and

3. that The IDEA Center's copyright be acknowledged.

We appreciate your request to use and interest in the materials of The IDEA Center.

Sincerely yours,

Ken Ryalls, Ph.D.
President

cc: Steve Benton, Ph.D.

Student Ratings Diagnostic Form (SRDF)

THE IDEA CENTER

SURVEY FORM - STUDENT REACTIONS TO INSTRUCTION AND COURSES

IMPORTANT! USE NO. 2 PENCIL ONLY

Proper Marks ● ● ● ● ● ● Improper Marks

Institution: Instructor:

Course Number: Time and Days Class Meets:

Your thoughtful answers to these questions will provide helpful information to your instructor.

Describe the frequency of your instructor's teaching procedures, using the following code:

1=Hardly Ever 2=Occasionally 3=Sometimes 4=Frequently 5=Almost Always

The Instructor:

1. ① ② ③ ④ ⑤ Displayed a personal interest in students and their learning
2. ① ② ③ ④ ⑤ Found ways to help students answer their own questions
3. ① ② ③ ④ ⑤ Scheduled course work (class activities, tests, projects) in ways which encouraged students to stay up-to-date in their work
4. ① ② ③ ④ ⑤ Demonstrated the importance and significance of the subject matter
5. ① ② ③ ④ ⑤ Formed "teams" or "discussion groups" to facilitate learning
6. ① ② ③ ④ ⑤ Made it clear how each topic fit into the course
7. ① ② ③ ④ ⑤ Explained the reasons for criticisms of students' academic performance
8. ① ② ③ ④ ⑤ Stimulated students to intellectual effort beyond that required by most courses
9. ① ② ③ ④ ⑤ Encouraged students to use multiple resources (e.g. data banks, library holdings, outside experts) to improve understanding
10. ① ② ③ ④ ⑤ Explained course material clearly and concisely
11. ① ② ③ ④ ⑤ Related course material to real life situations
12. ① ② ③ ④ ⑤ Gave tests, projects, etc. that covered the most important points of the course
13. ① ② ③ ④ ⑤ Introduced stimulating ideas about the subject
14. ① ② ③ ④ ⑤ Involved students in "hands on" projects such as research, case studies, or "real life" activities
15. ① ② ③ ④ ⑤ Inspired students to set and achieve goals which really challenged them
16. ① ② ③ ④ ⑤ Asked students to share ideas and experiences with others whose backgrounds and viewpoints differ from their own
17. ① ② ③ ④ ⑤ Provided timely and frequent feedback on tests, reports, projects, etc. to help students improve
18. ① ② ③ ④ ⑤ Asked students to help each other understand ideas or concepts
19. ① ② ③ ④ ⑤ Gave projects, tests, or assignments that required original or creative thinking
20. ① ② ③ ④ ⑤ Encouraged student-faculty interaction outside of class (office visits, phone calls, e-mail, etc.)

Twelve possible learning objectives are listed below, not all of which will be relevant in this class. Describe the amount of progress you made on each (even those not pursued in this class) by using the following scale:

1-No apparent progress
2-Slight progress; I made small gains on this objective.
3-Moderate progress; I made some gains on this objective.
4-Substantial progress; I made large gains on this objective.
5-Exceptional progress; I made outstanding gains on this objective.

Progress on:

21. ① ② ③ ④ ⑤ Gaining factual knowledge (terminology, classifications, methods, trends)
22. ① ② ③ ④ ⑤ Learning fundamental principles, generalizations, or theories
23. ① ② ③ ④ ⑤ Learning to *apply* course material (to improve thinking, problem solving, and decisions)
24. ① ② ③ ④ ⑤ Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course
25. ① ② ③ ④ ⑤ Acquiring skills in working with others as a member of a team
26. ① ② ③ ④ ⑤ Developing creative capacities (writing, inventing, designing, performing in art, music, drama, etc.)
27. ① ② ③ ④ ⑤ Gaining a broader understanding and appreciation of intellectual/cultural activity (music, science, literature, etc.)
28. ① ② ③ ④ ⑤ Developing skill in expressing myself orally or in writing
29. ① ② ③ ④ ⑤ Learning how to find and use resources for answering questions or solving problems
30. ① ② ③ ④ ⑤ Developing a clearer understanding of, and commitment to, personal values
31. ① ② ③ ④ ⑤ Learning to *analyze* and *critically evaluate* ideas, arguments, and points of view
32. ① ② ③ ④ ⑤ Acquiring an interest in learning more by asking my own questions and seeking answers

Continued on back page

On the next three items, compare this course with others you have taken at this institution, using the following code:

1=Much Less than Most Courses | 2=Less than Most Courses | 3=About Average | 4=More than Most Courses | 5=Much More than Most Courses

The Course:

33. ① ② ③ ④ ⑤ Amount of reading
34. ① ② ③ ④ ⑤ Amount of work in other (non-reading) assignments
35. ① ② ③ ④ ⑤ Difficulty of subject matter

Describe your attitudes and behavior in this course, using the following code:

1=Definitely False | 2=More False Than True | 3=In Between | 4=More True Than False | 5=Definitely True

36. ① ② ③ ④ ⑤ I had a strong desire to take this course.
37. ① ② ③ ④ ⑤ I worked harder on this course than on most courses I have taken.
38. ① ② ③ ④ ⑤ I really wanted to take a course from this instructor.
39. ① ② ③ ④ ⑤ I really wanted to take this course regardless of who taught it.
40. ① ② ③ ④ ⑤ As a result of taking this course, I have more positive feelings toward this field of study.
41. ① ② ③ ④ ⑤ Overall, I rate this instructor an excellent teacher.
42. ① ② ③ ④ ⑤ Overall, I rate this course as excellent.

For the following items, blacken the space which best corresponds to your judgment:

1=Definitely False | 2=More False Than True | 3=In Between | 4=More True Than False | 5=Definitely True

43. ① ② ③ ④ ⑤ As a rule, I put forth more effort than other students on academic work.
44. ① ② ③ ④ ⑤ The instructor used a variety of methods--not only tests--to evaluate student progress on course objectives.
45. ① ② ③ ④ ⑤ The instructor expected students to take their share of responsibility for learning.
46. ① ② ③ ④ ⑤ The instructor had high achievement standards in this class.
47. ① ② ③ ④ ⑤ The instructor used educational technology (e.g., Internet, e-mail, computer exercises, multi-media presentations, etc.) to promote learning.

EXTRA QUESTIONS

If your instructor has extra questions, answer them in the space designated below (questions 48-67):

48. ① ② ③ ④ ⑤
49. ① ② ③ ④ ⑤
50. ① ② ③ ④ ⑤
51. ① ② ③ ④ ⑤
52. ① ② ③ ④ ⑤
53. ① ② ③ ④ ⑤
54. ① ② ③ ④ ⑤
55. ① ② ③ ④ ⑤
56. ① ② ③ ④ ⑤
57. ① ② ③ ④ ⑤
58. ① ② ③ ④ ⑤
59. ① ② ③ ④ ⑤
60. ① ② ③ ④ ⑤
61. ① ② ③ ④ ⑤
62. ① ② ③ ④ ⑤
63. ① ② ③ ④ ⑤
64. ① ② ③ ④ ⑤
65. ① ② ③ ④ ⑤
66. ① ② ③ ④ ⑤
67. ① ② ③ ④ ⑤

Use the space below for comments (unless otherwise directed).
Note: Your written comments may be returned to the instructor, You may want to PRINT to protect your anonymity.

Comments: ____________________

TF5903 (08/08) 0 9 8 7 6 5 Printed in U.S.A.

Student Ratings Short Form

SHORT FORM - STUDENT REACTIONS TO INSTRUCTION AND COURSES

IDEA CENTER

Institution:	Instructor:
Course Number:	Time and Days Class Meets:

IMPORTANT! USE NO. 2 PENCIL ONLY

Proper Marks ●●●●●●

Improper Marks

Twelve possible learning objectives are listed below, not all of which will be relevant in this class. Describe the amount of progress you made on each (even those not pursued in this class) by using the following scale:

1-No apparent progress
2-Slight progress; I made small gains on this objective.
3-Moderate progress; I made some gains on this objective.
4-Substantial progress; I made large gains on this objective.
5-Exceptional progress; I made outstanding gains on this objective.

Progress on:

1. Gaining factual knowledge (terminology, classifications, methods, trends)
2. Learning fundamental principles, generalizations, or theories
3. Learning to *apply* course material (to improve thinking, problem solving, and decisions)
4. Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course
5. Acquiring skills in working with others as a member of a team
6. Developing creative capacities (writing, inventing, designing, performing in art, music, drama, etc.)
7. Gaining a broader understanding and appreciation of intellectual/cultural activity (music, science, literature, etc.)
8. Developing skill in expressing myself orally or in writing
9. Learning how to find and use resources for answering questions or solving problems
10. Developing a clearer understanding of, and commitment to, personal values
11. Learning to *analyze* and *critically evaluate* ideas, arguments, and points of view
12. Acquiring an interest in learning more by asking my own questions and seeking answers

For the remaining questions, use the following code:

1=Definitely False	2=More False Than True	3=In Between	4=More True Than False	5=Definitely True

13. As a rule, I put forth more effort than other students on academic work.
14. My background prepared me well for this course's requirements.
15. I really wanted to take this course regardless of who taught it.
16. As a result of taking this course, I have more positive feelings toward this field of study.
17. Overall, I rate this instructor an excellent teacher.
18. Overall, I rate this course as excellent.

EXTRA QUESTIONS

If your instructor has extra questions, answer them in the space designated below (questions 19-38).

19.–38.

Use the space provided on the back of this form for your comments.

Continue on back page

Comments:

DO NOT

WRITE

IN THE

SHADED

AREA

TF5902 (08/08) 0 9 8 7 6 5

Printed in U.S.A.

Faculty Information Form (FIF)

Faculty Information Form

IDEA CENTER

See Directions to Faculty:
www.theideacenter.org/directions

IMPORTANT!

USE NO. 2 PENCIL ONLY

Proper Marks

Improper Marks

Institution: ______________________

Instructor: ______________________

Course Number: ______________________

Time and Days Class Meets: ______________________

Last Name (Up to 11 letters) | **Init.**

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Objectives: Using the scale provided, identify the relevance of each of the twelve objectives to this course. As a general rule, prioritize what you want students to learn by selecting no more than 3-5 objectives as either Important or Essential. The weighting system used to generate the IDEA report weighs Essential objectives "2," Important objectives "1," and Minor objectives "0."
(Scale - M = Minor or No Importance, I = Important, E = Essential)

M I E

1. Gaining factual knowledge (terminology, classifications, methods, trends)
2. Learning fundamental principles, generalizations, or theories
3. Learning to *apply* course material (to improve thinking, problem solving, and decisions)
4. Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course
5. Acquiring skills in working with others as a member of a team
6. Developing creative capacities (writing, inventing, designing, performing in art, music, drama, etc.)
7. Gaining a broader understanding and appreciation of intellectual/cultural activity (music, science, literature, etc.)
8. Developing skill in expressing oneself orally or in writing
9. Learning how to find and use resources for answering questions or solving problems
10. Developing a clearer understanding of, and commitment to, personal values
11. Learning to *analyze* and *critically evaluate* ideas, arguments, and points of view
12. Acquiring an interest in learning more by asking questions and seeking answers

Days Class Meets: Mon, Tues, Wed, Thu, Fri, Sat, Sun

Discipline Code (0–9)

Time Class Begins (0–9)

Course Number (0–9)

Number Enrolled (0–9)

Local Codes: A B C D E F G H (0–9)

Contextual Questions (Research Purposes):

The IDEA Center will conduct research on these optional questions in order to improve the interpretation of student ratings.

1. Which of the following represents the <u>primary</u> approach to this course? (Mark only one)

1 = Lecture
2 = Discussion/recitation
3 = Seminar
4 = Skill/activity
5 = Laboratory
6 = Field Experience
7 = Studio
8 = Multi-Media
9 = Practicum/clinic
0 = Other

2. If multiple approaches are used, which <u>one</u> represents the <u>secondary approach</u>? (Mark only one)

1 = Lecture
2 = Discussion/recitation
3 = Seminar
4 = Skill/activity
5 = Laboratory
6 = Field Experience
7 = Studio
8 = Multi-Media
9 = Practicum/clinic
0 = Other

3. Describe this course in terms of its requirements with respect to the features listed below. Use the following code to make your responses:
N = None (or little) required
S = Some required
M = Much required

N S M

A. Writing
B. Oral communication
C. Computer applications
D. Group work
E. Mathematical/quantitative work
F. Critical thinking
G. Creative/artistic/design endeavor
H. Reading
I. Memorization

TF5901 (08/11) 0 9 8 7 6 5 4 3 2 1

Printed in U.S.A.

Continue on back page

Contextual Questions Continued:

4. Rate each of the circumstances listed below, using the following code to respond:

P = Had a positive impact on learning
I = Neither a positive nor a negative impact
N = Had a negative impact on learning
? = Can't judge

P	I	N	?	
◯	◯	◯	◯	A. Physical facilities and/or equipment
◯	◯	◯	◯	B. Your previous experience in teaching this course
◯	◯	◯	◯	C. Substantial changes in teaching approach, course assignments, content, etc.
◯	◯	◯	◯	D. Your desire to teach this course
◯	◯	◯	◯	E. Your control over course management decisions (objectives, texts, exams, etc.)
◯	◯	◯	◯	F. Students' level of preparation for taking the course
◯	◯	◯	◯	G. Students' level of enthusiasm for the course
◯	◯	◯	◯	H. Students' level of effort to learn
◯	◯	◯	◯	I. Technical/instructional support

5. Please identify the principal type of student enrolling in this course (Mark only one)

① = First-year students/sophomores seeking to meet a "general education" or "distribution" requirement
② = First-year students/sophomores seeking to develop background needed for their intended specialization
③ = Upper level non-majors taking the course as a "general education" or "distribution" requirement
④ = Upper level majors (in this or a related field of study) seeking competence or expertise in their academic/professional specialty
⑤ = Graduate or professional school students
⑥ = Combination of two or more of the above types

6. Is this class:

a. Team taught? ◯ Yes ◯ No
b. Taught through distance learning? ◯ Yes ◯ No

Discipline Codes (Modified CIP Codes)

0100 Agricultural Business and Production
0200 Agricultural Sciences
0300 Conservation and Renewable Natural Resources
0400 Architecture and Related Programs
0500 Area Ethnic and Cultural Studies
5007 Art (Painting, Drawing, Sculpture)
3201 Basic Skills
2600 Biological Sciences/Life Sciences
5201 Business, General
5202 Business Administration and Management
5203 Business - Accounting
5208 Business - Finance
5212 Business Information and Data Processing Services
5214 Business - Marketing
4005 Chemistry
0900 Communications
1100 Computer and Information Sciences
4301 Criminal Justice and Corrections
1205 Culinary Arts and Related Services
1103 Data Processing Technology (2-year program)
5004 Design and Applied Arts
9901 Developmental Math
9902 Developmental Reading
9903 Developmental Writing
9904 Developmental Natural Sciences
4506 Economics
1300 Education
1400 Engineering
1500 Engineering-Related Technologies
9910 English as Second Language
2301 English Language and Literature
5000 Fine and Applied Arts (EXCEPT Art, Music, and Design and Applied Arts)
1600 Foreign Languages and Literatures
3105 Health and Physical Education/Fitness
5100 Health Professions and Related Sciences (EXCEPT Nursing)
5199 Health Professions and Related Sciences (2-year program)
4508 History
1900 Human Sciences/Family and Consumer Sciences
2400 Liberal Arts & Sciences, General Studies and Humanities
2200 General Legal Studies (Undergraduate)
2500 Library Science
2700 Mathematics and Statistics
5009 Music (Performing, Composing, Theory)
5116 Nursing
3100 Parks, Recreation, Leisure, and Fitness Studies
3801 Philosophy
4000 Physical Science (EXCEPT Physics and Chemistry)
4008 Physics
4510 Political Science and Government
4200 Psychology
4400 Public Administration and Services (EXCEPT Social Work)
3900 Religion and Theological Studies
4500 Social Sciences (EXCEPT Economics, History, Political Science, and Sociology)
4407 Social Work and Service
4511 Sociology
2310 Speech and Rhetorical Studies

Vocational/Technical Programs
(see Website: Department codes 4600-4900)

9900 Other (to be used when none of the above codes apply)

To see an expanded list of discipline codes go to: www.theideacenter.org/DisciplineCodes

Appendix B - List of STEM CIP Codes

2010 CIP Code	CIP Title
1.0308	Agroecology and Sustainable Agriculture
1.0901	Animal Sciences, General
1.0902	Agricultural Animal Breeding
1.0903	Animal Health
1.0904	Animal Nutrition
1.0905	Dairy Science
1.0906	Livestock Management
1.0907	Poultry Science
1.0999	Animal Sciences, Other
1.1001	Food Science
1.1002	Food Technology and Processing
1.1099	Food Science and Technology, Other
1.1101	Plant Sciences, General
1.1102	Agronomy and Crop Science
1.1103	Horticultural Science
1.1104	Agricultural and Horticultural Plant Breeding
1.1105	Plant Protection and Integrated Pest Management
1.1106	Range Science and Management
1.1199	Plant Sciences, Other
1.1201	Soil Science and Agronomy, General
1.1202	Soil Chemistry and Physics
1.1203	Soil Microbiology
1.1299	Soil Sciences, Other
3.0101	Natural Resources/Conservation, General
3.0103	Environmental Studies
3.0104	Environmental Science
3.0199	Natural Resources Conservation and Research, Other
3.0205	Water, Wetlands, and Marine Resources Management
3.0502	Forest Sciences and Biology
3.0508	Urban Forestry

2010 CIP Code	CIP Title
3.0509	Wood Science and Wood Products/Pulp and Paper Technology
3.0601	Wildlife, Fish and Wildlands Science and Management
4.0902	Architectural and Building Sciences/Technology
9.0702	Digital Communication and Media/Multimedia
10.0304	Animation, Interactive Technology, Video Graphics and Special Effects
11.0101	Computer and Information Sciences, General
11.0102	Artificial Intelligence
11.0103	Information Technology
11.0104	Informatics
11.0199	Computer and Information Sciences, Other
11.0201	Computer Programming/Programmer, General
11.0202	Computer Programming, Specific Applications
11.0203	Computer Programming, Vendor/Product Certification
11.0299	Computer Programming, Other
11.0301	Data Processing and Data Processing Technology/Technician
11.0401	Information Science/Studies
11.0501	Computer Systems Analysis/Analyst
11.0701	Computer Science
11.0801	Web Page, Digital/Multimedia and Information Resources Design
11.0802	Data Modeling/Warehousing and Database Administration
11.0803	Computer Graphics
11.0804	Modeling, Virtual Environments and Simulation
11.0899	Computer Software and Media Applications, Other
11.0901	Computer Systems Networking and Telecommunications
11.1001	Network and System Administration/Administrator
11.1002	System, Networking, and LAN/WAN Management/Manager
11.1003	Computer and Information Systems Security/Information Assurance
11.1004	Web/Multimedia Management and Webmaster
11.1005	Information Technology Project Management

2010 CIP Code	CIP Title
11.1006	Computer Support Specialist
11.1099	Computer/Information Technology Services Administration and Management, Other
13.0501	Educational/Instructional Technology
13.0601	Educational Evaluation and Research
13.0603	Educational Statistics and Research Methods
14.0101	Engineering, General
14.0102	Pre-Engineering
14.0201	Aerospace, Aeronautical and Astronautical/Space Engineering
14.0301	Agricultural Engineering
14.0401	Architectural Engineering
14.0501	Bioengineering and Biomedical Engineering
14.0601	Ceramic Sciences and Engineering
14.0701	Chemical Engineering
14.0702	Chemical and Biomolecular Engineering
14.0799	Chemical Engineering, Other
14.0801	Civil Engineering, General
14.0802	Geotechnical and Geoenvironmental Engineering
14.0803	Structural Engineering
14.0804	Transportation and Highway Engineering
14.0805	Water Resources Engineering
14.0899	Civil Engineering, Other
14.0901	Computer Engineering, General
14.0902	Computer Hardware Engineering
14.0903	Computer Software Engineering
14.0999	Computer Engineering, Other
14.1001	Electrical and Electronics Engineering
14.1003	Laser and Optical Engineering
14.1004	Telecommunications Engineering
14.1099	Electrical, Electronics and Communications Engineering, Other

2010 CIP Code	CIP Title
14.1101	Engineering Mechanics
14.1201	Engineering Physics/Applied Physics
14.1301	Engineering Science
14.1401	Environmental/Environmental Health Engineering
14.1801	Materials Engineering
14.1901	Mechanical Engineering
14.2001	Metallurgical Engineering
14.2101	Mining and Mineral Engineering
14.2201	Naval Architecture and Marine Engineering
14.2301	Nuclear Engineering
14.2401	Ocean Engineering
14.2501	Petroleum Engineering
14.2701	Systems Engineering
14.2801	Textile Sciences and Engineering
14.3201	Polymer/Plastics Engineering
14.3301	Construction Engineering
14.3401	Forest Engineering
14.3501	Industrial Engineering
14.3601	Manufacturing Engineering
14.3701	Operations Research
14.3801	Surveying Engineering
14.3901	Geological/Geophysical Engineering
14.4001	Paper Science and Engineering
14.4101	Electromechanical Engineering
14.4201	Mechatronics, Robotics, and Automation Engineering
14.4301	Biochemical Engineering
14.4401	Engineering Chemistry
14.4501	Biological/Biosystems Engineering

2010 CIP Code	CIP Title
14.9999	Engineering, Other
15.0000	Engineering Technology, General
15.0101	Architectural Engineering Technology/Technician
15.0201	Civil Engineering Technology/Technician
15.0303	Electrical, Electronic and Communications Engineering Technology/Technician
15.0304	Laser and Optical Technology/Technician
15.0305	Telecommunications Technology/Technician
15.0306	Integrated Circuit Design
15.0399	Electrical and Electronic Engineering Technologies/Technicians, Other
15.0401	Biomedical Technology/Technician
15.0403	Electromechanical Technology/Electromechanical Engineering Technology
15.0404	Instrumentation Technology/Technician
15.0405	Robotics Technology/Technician
15.0406	Automation Engineer Technology/Technician
15.0499	Electromechanical and Instrumentation and Maintenance Technologies/Technicians, Other
15.0501	Heating, Ventilation, Air Conditioning and Refrigeration Engineering Technology/Technician
15.0503	Energy Management and Systems Technology/Technician
15.0505	Solar Energy Technology/Technician
15.0506	Water Quality and Wastewater Treatment Management and Recycling Technology/Technician
15.0507	Environmental Engineering Technology/Environmental Technology
15.0508	Hazardous Materials Management and Waste Technology/Technician
15.0599	Environmental Control Technologies/Technicians, Other
15.0607	Plastics and Polymer Engineering Technology/Technician
15.0611	Metallurgical Technology/Technician
15.0612	Industrial Technology/Technician
15.0613	Manufacturing Engineering Technology/Technician
15.0614	Welding Engineering Technology/Technician
15.0615	Chemical Engineering Technology/Technician
15.0616	Semiconductor Manufacturing Technology

2010 CIP Code	CIP Title
15.0699	Industrial Production Technologies/Technicians, Other
15.0701	Occupational Safety and Health Technology/Technician
15.0702	Quality Control Technology/Technician
15.0703	Industrial Safety Technology/Technician
15.0704	Hazardous Materials Information Systems Technology/Technician
15.0799	Quality Control and Safety Technologies/Technicians, Other
15.0801	Aeronautical/Aerospace Engineering Technology/Technician
15.0803	Automotive Engineering Technology/Technician
15.0805	Mechanical Engineering/Mechanical Technology/Technician
15.0899	Mechanical Engineering Related Technologies/Technicians, Other
15.0901	Mining Technology/Technician
15.0903	Petroleum Technology/Technician
15.0999	Mining and Petroleum Technologies/Technicians, Other
15.1001	Construction Engineering Technology/Technician
15.1102	Surveying Technology/Surveying
15.1103	Hydraulics and Fluid Power Technology/Technician
15.1199	Engineering-Related Technologies, Other
15.1201	Computer Engineering Technology/Technician
15.1202	Computer Technology/Computer Systems Technology
15.1203	Computer Hardware Technology/Technician
15.1204	Computer Software Technology/Technician
15.1299	Computer Engineering Technologies/Technicians, Other
15.1301	Drafting and Design Technology/Technician, General
15.1302	CAD/CADD Drafting and/or Design Technology/Technician
15.1303	Architectural Drafting and Architectural CAD/CADD
15.1304	Civil Drafting and Civil Engineering CAD/CADD
15.1305	Electrical/Electronics Drafting and Electrical/Electronics CAD/CADD
15.1306	Mechanical Drafting and Mechanical Drafting CAD/CADD
15.1399	Drafting/Design Engineering Technologies/Technicians, Other

2010 CIP Code	CIP Title
15.1401	Nuclear Engineering Technology/Technician
15.1501	Engineering/Industrial Management
15.1502	Engineering Design
15.1503	Packaging Science
15.1599	Engineering-Related Fields, Other
15.1601	Nanotechnology
15.9999	Engineering Technologies and Engineering-Related Fields, Other
26.0101	Biology/Biological Sciences, General
26.0102	Biomedical Sciences, General
26.0202	Biochemistry
26.0203	Biophysics
26.0204	Molecular Biology
26.0205	Molecular Biochemistry
26.0206	Molecular Biophysics
26.0207	Structural Biology
26.0208	Photobiology
26.0209	Radiation Biology/Radiobiology
26.0210	Biochemistry and Molecular Biology
26.0299	Biochemistry, Biophysics and Molecular Biology, Other
26.0301	Botany/Plant Biology
26.0305	Plant Pathology/Phytopathology
26.0307	Plant Physiology
26.0308	Plant Molecular Biology
26.0399	Botany/Plant Biology, Other
26.0401	Cell/Cellular Biology and Histology
26.0403	Anatomy
26.0404	Developmental Biology and Embryology
26.0406	Cell/Cellular and Molecular Biology
26.0407	Cell Biology and Anatomy

2010 CIP Code	CIP Title
26.0499	Cell/Cellular Biology and Anatomical Sciences, Other
26.0502	Microbiology, General
26.0503	Medical Microbiology and Bacteriology
26.0504	Virology
26.0505	Parasitology
26.0506	Mycology
26.0507	Immunology
26.0508	Microbiology and Immunology
26.0599	Microbiological Sciences and Immunology, Other
26.0701	Zoology/Animal Biology
26.0702	Entomology
26.0707	Animal Physiology
26.0708	Animal Behavior and Ethology
26.0709	Wildlife Biology
26.0799	Zoology/Animal Biology, Other
26.0801	Genetics, General
26.0802	Molecular Genetics
26.0803	Microbial and Eukaryotic Genetics
26.0804	Animal Genetics
26.0805	Plant Genetics
26.0806	Human/Medical Genetics
26.0807	Genome Sciences/Genomics
26.0899	Genetics, Other
26.0901	Physiology, General
26.0902	Molecular Physiology
26.0903	Cell Physiology
26.0904	Endocrinology
26.0905	Reproductive Biology
26.0907	Cardiovascular Science

2010 CIP Code	CIP Title
26.0908	Exercise Physiology
26.0909	Vision Science/Physiological Optics
26.0910	Pathology/Experimental Pathology
26.0911	Oncology and Cancer Biology
26.0912	Aerospace Physiology and Medicine
26.0999	Physiology, Pathology, and Related Sciences, Other
26.1001	Pharmacology
26.1002	Molecular Pharmacology
26.1003	Neuropharmacology
26.1004	Toxicology
26.1005	Molecular Toxicology
26.1006	Environmental Toxicology
26.1007	Pharmacology and Toxicology
26.1099	Pharmacology and Toxicology, Other
26.1101	Biometry/Biometrics
26.1102	Biostatistics
26.1103	Bioinformatics
26.1104	Computational Biology
26.1199	Biomathematics, Bioinformatics, and Computational Biology, Other
26.1201	Biotechnology
26.1301	Ecology
26.1302	Marine Biology and Biological Oceanography
26.1303	Evolutionary Biology
26.1304	Aquatic Biology/Limnology
26.1305	Environmental Biology
26.1306	Population Biology
26.1307	Conservation Biology
26.1308	Systematic Biology/Biological Systematics
26.1309	Epidemiology

2010 CIP Code	CIP Title
26.1310	Ecology and Evolutionary Biology
26.1399	Ecology, Evolution, Systematics and Population Biology, Other
26.1401	Molecular Medicine
26.1501	Neuroscience
26.1502	Neuroanatomy
26.1503	Neurobiology and Anatomy
26.1504	Neurobiology and Behavior
26.1599	Neurobiology and Neurosciences, Other
26.9999	Biological and Biomedical Sciences, Other
27.0101	Mathematics, General
27.0102	Algebra and Number Theory
27.0103	Analysis and Functional Analysis
27.0104	Geometry/Geometric Analysis
27.0105	Topology and Foundations
27.0199	Mathematics, Other
27.0301	Applied Mathematics, General
27.0303	Computational Mathematics
27.0304	Computational and Applied Mathematics
27.0305	Financial Mathematics
27.0306	Mathematical Biology
27.0399	Applied Mathematics, Other
27.0501	Statistics, General
27.0502	Mathematical Statistics and Probability
27.0503	Mathematics and Statistics
27.0599	Statistics, Other
27.9999	Mathematics and Statistics, Other
28.0501	Air Science/Airpower Studies
28.0502	Air and Space Operational Art and Science
28.0505	Naval Science and Operational Studies

2010 CIP Code	CIP Title
29.0201	Intelligence, General
29.0202	Strategic Intelligence
29.0203	Signal/Geospatial Intelligence
29.0204	Command & Control (C3, C4I) Systems and Operations
29.0205	Information Operations/Joint Information Operations
29.0206	Information/Psychological Warfare and Military Media Relations
29.0207	Cyber/Electronic Operations and Warfare
29.0299	Intelligence, Command Control and Information Operations, Other
29.0301	Combat Systems Engineering
29.0302	Directed Energy Systems
29.0303	Engineering Acoustics
29.0304	Low-Observables and Stealth Technology
29.0305	Space Systems Operations
29.0306	Operational Oceanography
29.0307	Undersea Warfare
29.0399	Military Applied Sciences, Other
29.0401	Aerospace Ground Equipment Technology
29.0402	Air and Space Operations Technology
29.0403	Aircraft Armament Systems Technology
29.0404	Explosive Ordinance/Bomb Disposal
29.0405	Joint Command/Task Force (C3, C4I) Systems
29.0406	Military Information Systems Technology
29.0407	Missile and Space Systems Technology
29.0408	Munitions Systems/Ordinance Technology
29.0409	Radar Communications and Systems Technology
29.0499	Military Systems and Maintenance Technology, Other
29.9999	Military Technologies and Applied Sciences, Other
30.0101	Biological and Physical Sciences
30.0601	Systems Science and Theory

2010 CIP Code	CIP Title
30.0801	Mathematics and Computer Science
30.1001	Biopsychology
30.1701	Behavioral Sciences
30.1801	Natural Sciences
30.1901	Nutrition Sciences
30.2501	Cognitive Science
30.2701	Human Biology
30.3001	Computational Science
30.3101	Human Computer Interaction
30.3201	Marine Sciences
30.3301	Sustainability Studies
40.0101	Physical Sciences
40.0201	Astronomy
40.0202	Astrophysics
40.0203	Planetary Astronomy and Science
40.0299	Astronomy and Astrophysics, Other
40.0401	Atmospheric Sciences and Meteorology, General
40.0402	Atmospheric Chemistry and Climatology
40.0403	Atmospheric Physics and Dynamics
40.0404	Meteorology
40.0499	Atmospheric Sciences and Meteorology, Other
40.0501	Chemistry, General
40.0502	Analytical Chemistry
40.0503	Inorganic Chemistry
40.0504	Organic Chemistry
40.0506	Physical Chemistry
40.0507	Polymer Chemistry
40.0508	Chemical Physics
40.0509	Environmental Chemistry

2010 CIP Code	CIP Title
40.0510	Forensic Chemistry
40.0511	Theoretical Chemistry
40.0599	Chemistry, Other
40.0601	Geology/Earth Science, General
40.0602	Geochemistry
40.0603	Geophysics and Seismology
40.0604	Paleontology
40.0605	Hydrology and Water Resources Science
40.0606	Geochemistry and Petrology
40.0607	Oceanography, Chemical and Physical
40.0699	Geological and Earth Sciences/Geosciences, Other
40.0801	Physics, General
40.0802	Atomic/Molecular Physics
40.0804	Elementary Particle Physics
40.0805	Plasma and High-Temperature Physics
40.0806	Nuclear Physics
40.0807	Optics/Optical Sciences
40.0808	Condensed Matter and Materials Physics
40.0809	Acoustics
40.0810	Theoretical and Mathematical Physics
40.0899	Physics, Other
40.1001	Materials Science
40.1002	Materials Chemistry
40.1099	Materials Sciences, Other
40.9999	Physical Sciences, Other
41.0000	Science Technologies/Technicians, General
41.0101	Biology Technician/Biotechnology Laboratory Technician
41.0204	Industrial Radiologic Technology/Technician
41.0205	Nuclear/Nuclear Power Technology/Technician

2010 CIP Code	CIP Title
41.0299	Nuclear and Industrial Radiologic Technologies/Technicians, Other
41.0301	Chemical Technology/Technician
41.0303	Chemical Process Technology
41.0399	Physical Science Technologies/Technicians, Other
41.9999	Science Technologies/Technicians, Other
42.2701	Cognitive Psychology and Psycholinguistics
42.2702	Comparative Psychology
42.2703	Developmental and Child Psychology
42.2704	Experimental Psychology
42.2705	Personality Psychology
42.2706	Physiological Psychology/Psychobiology
42.2707	Social Psychology
42.2708	Psychometrics and Quantitative Psychology
42.2709	Psychopharmacology
42.2799	Research and Experimental Psychology, Other
43.0106	Forensic Science and Technology
43.0116	Cyber/Computer Forensics and Counterterrorism
45.0301	Archeology
45.0603	Econometrics and Quantitative Economics
45.0702	Geographic Information Science and Cartography
49.0101	Aeronautics/Aviation/Aerospace Science and Technology, General
51.1002	Cytotechnology/Cytotechnologist
51.1005	Clinical Laboratory Science/Medical Technology/Technologist
51.1401	Medical Scientist
51.2003	Pharmaceutics and Drug Design
51.2004	Medicinal and Pharmaceutical Chemistry
51.2005	Natural Products Chemistry and Pharmacognosy
51.2006	Clinical and Industrial Drug Development
51.2007	Pharmacoeconomics/Pharmaceutical Economics

2010 CIP Code	CIP Title
51.2009	Industrial and Physical Pharmacy and Cosmetic Sciences
51.2010	Pharmaceutical Sciences
51.2202	Environmental Health
51.2205	Health/Medical Physics
51.2502	Veterinary Anatomy
51.2503	Veterinary Physiology
51.2504	Veterinary Microbiology and Immunobiology
51.2505	Veterinary Pathology and Pathobiology
51.2506	Veterinary Toxicology and Pharmacology
51.2510	Veterinary Preventive Medicine, Epidemiology, and Public Health
51.2511	Veterinary Infectious Diseases
51.2706	Medical Informatics
52.1301	Management Science
52.1302	Business Statistics
52.1304	Actuarial Science
52.1399	Management Sciences and Quantitative Methods, Other

Appendix C - STATISTICAL TABLES

Research Question 1

Dependent Variable: Overall I rate this instructor an excellent teacher (non-doctoral)

Discipline	Course Type	Student Type	N	Mean	Std. Deviation
STEM	Lecture /Discussion /Seminar	Lower Level, General Education	4,305	4.14	0.63
		Lower Level, Specialization	4,948	4.14	0.67
		Upper Level, General Education	1,157	4.08	0.68
		Upper Level, Specialization	2,942	4.22	0.66
		Graduate/Professional	140	4.26	0.51
		Total	13,492	4.15	0.66
	Laboratory & Studio	Lower Level, General Education	739	4.16	0.57
		Lower Level, Specialization	1,374	4.31	0.56
		Upper Level, General Education	130	4.06	0.69
		Upper Level, Specialization	453	4.23	0.64
		Graduate/Professional	5	4.36	0.39
		Total	2,701	4.24	0.59
	Activity /Field /Media /Practicum	Lower Level, General Education	743	4.12	0.65
		Lower Level, Specialization	519	4.14	0.71
		Upper Level, General Education	99	4.18	0.65
		Upper Level, Specialization	257	4.24	0.66
		Graduate/Professional	15	3.94	0.83
		Total	1,633	4.15	0.68
	Total	Lower Level, General Education	5,787	4.14	0.62
		Lower Level, Specialization	6,841	4.18	0.66
		Upper Level, General Education	1,386	4.08	0.68
		Upper Level, Specialization	3,652	4.22	0.66
		Graduate/Professional	160	4.23	0.55
		Total	17,826	4.17	0.65
Education	Lecture /Discussion /Seminar	Lower Level, General Education	375	4.25	0.62
		Lower Level, Specialization	1,122	4.39	0.52
		Upper Level, General Education	152	4.24	0.64
		Upper Level, Specialization	2,165	4.32	0.63
		Graduate/Professional	1,181	4.44	0.55
		Total	4,995	4.36	0.59
	Laboratory & Studio	Lower Level, General Education	18	4.45	0.25
		Lower Level, Specialization	28	4.22	0.60
		Upper Level, General Education	8	4.17	0.72

		Upper Level, Specialization	63	4.24	0.50
		Graduate/Professional	8	4.55	0.37
		Total	125	4.28	0.51
	Activity /Field /Media /Practicum	Lower Level, General Education	116	4.17	0.66
		Lower Level, Specialization	291	4.35	0.62
		Upper Level, General Education	85	4.42	0.56
		Upper Level, Specialization	676	4.27	0.64
		Graduate/Professional	351	4.22	0.60
		Total	1,519	4.28	0.63
	Total	Lower Level, General Education	509	4.24	0.62
		Lower Level, Specialization	1,441	4.38	0.54
		Upper Level, General Education	245	4.30	0.62
		Upper Level, Specialization	2,904	4.31	0.63
		Graduate/Professional	1,540	4.39	0.57
		Total	6,639	4.34	0.60
Business	Lecture /Discussion /Seminar	Lower Level, General Education	873	4.08	0.63
		Lower Level, Specialization	2,026	4.23	0.57
		Upper Level, General Education	829	4.09	0.68
		Upper Level, Specialization	4,818	4.23	0.63
		Graduate/Professional	869	4.27	0.60
		Total	9,415	4.21	0.62
	Laboratory& Studio	Lower Level, General Education	58	3.85	0.52
		Lower Level, Specialization	32	4.04	0.59
		Upper Level, General Education	5	4.13	0.60
		Upper Level, Specialization	28	4.15	0.60
		Graduate/Professional	29	4.32	0.55
		Total	152	4.04	0.58
	Activity /Field /Media /Practicum	Lower Level, General Education	148	4.17	0.55
		Lower Level, Specialization	294	4.22	0.57
		Upper Level, General Education	81	4.14	0.52
		Upper Level, Specialization	536	4.17	0.63
		Graduate/Professional	113	4.13	0.72
		Total	1,172	4.18	0.61
	Total	Lower Level, General Education	1,079	4.08	0.62
		Lower Level, Specialization	2,352	4.23	0.57
		Upper Level, General Education	915	4.09	0.67
		Upper Level, Specialization	5,382	4.22	0.63
		Graduate/Professional	1,011	4.25	0.61
		Total	10,739	4.20	0.62
Liberal Arts		Lower Level, General Education	3,508	4.25	0.58

		Lower Level, Specialization	86	4.40	0.51
	Lecture /Discussion /Seminar	Upper Level, General Education	851	4.29	0.54
		Upper Level, Specialization	68	4.38	0.60
		Graduate/Professional	35	4.61	0.52
		Total	4,548	4.26	0.58
	Laboratory& Studio	Lower Level, General Education	19	4.52	0.31
		Lower Level, Specialization	5	4.56	0.28
		Upper Level, General Education	3	4.36	0.13
		Upper Level, Specialization	3	4.50	0.20
		Graduate/Professional			
		Total	30	4.51	0.28
	Activity /Field /Media /Practicum	Lower Level, General Education	446	4.26	0.55
		Lower Level, Specialization	19	4.26	0.79
		Upper Level, General Education	44	4.14	0.65
		Upper Level, Specialization	13	4.29	0.67
		Graduate/Professional	2	3.79	0.73
		Total	524	4.25	0.57
	Total	Lower Level, General Education	3,973	4.25	0.58
		Lower Level, Specialization	110	4.39	0.56
		Upper Level, General Education	898	4.28	0.54
		Upper Level, Specialization	84	4.37	0.60
		Graduate/Professional	37	4.56	0.55
		Total	5,102	4.26	0.57
Other	Lecture /Discussion /Seminar	Lower Level, General Education	18,237	4.28	0.54
		Lower Level, Specialization	8,437	4.31	0.59
		Upper Level, General Education	4,736	4.27	0.56
		Upper Level, Specialization	12,529	4.33	0.58
		Graduate/Professional	1,848	4.33	0.59
		Total	45,787	4.30	0.57
	Laboratory & Studio	Lower Level, General Education	966	4.31	0.54
		Lower Level, Specialization	2,084	4.39	0.56
		Upper Level, General Education	199	4.36	0.57
		Upper Level, Specialization	1,305	4.37	0.58
		Graduate/Professional	53	4.27	0.64
		Total	4,607	4.37	0.56
	Activity /Field /Media /Practicum	Lower Level, General Education	4,220	4.31	0.54
		Lower Level, Specialization	2,648	4.32	0.60
		Upper Level, General Education	792	4.31	0.59
		Upper Level, Specialization	2,203	4.32	0.60
		Graduate/Professional	317	4.26	0.62

		Total	10,180	4.31	0.57
	Total	Lower Level, General Education	23,423	4.28	0.54
		Lower Level, Specialization	13,169	4.32	0.58
		Upper Level, General Education	5,727	4.28	0.56
		Upper Level, Specialization	16,037	4.34	0.58
		Graduate/Professional	2,218	4.32	0.59
		Total	60,574	4.31	0.57
Total	Lecture /Discussion /Seminar	Lower Level, General Education	27,298	4.24	0.57
		Lower Level, Specialization	16,619	4.26	0.61
		Upper Level, General Education	7,725	4.22	0.60
		Upper Level, Specialization	22,522	4.30	0.61
		Graduate/Professional	4,073	4.35	0.58
		Total	78,237	4.27	0.59
	Laboratory & Studio	Lower Level, General Education	1,800	4.24	0.56
		Lower Level, Specialization	3,523	4.35	0.56
		Upper Level, General Education	345	4.24	0.63
		Upper Level, Specialization	1,852	4.33	0.59
		Graduate/Professional	95	4.31	0.58
		Total	7,615	4.31	0.58
	Activity /Field /Media /Practicum	Lower Level, General Education	5,673	4.27	0.57
		Lower Level, Specialization	3,771	4.29	0.62
		Upper Level, General Education	1,101	4.29	0.60
		Upper Level, Specialization	3,685	4.28	0.62
		Graduate/Professional	798	4.22	0.63
		Total	15,028	4.28	0.60
	Total	Lower Level, General Education	34,771	4.25	0.57
		Lower Level, Specialization	23,913	4.28	0.61
		Upper Level, General Education	9,171	4.23	0.60
		Upper Level, Specialization	28,059	4.30	0.61
		Graduate/Professional	4,966	4.33	0.59
		Total	100,880	4.27	0.59

Research Question 2

Dependent Variable: Overall I rate this instructor an excellent teacher (Doctoral)

Discipline	Course Type	Student Type	N	Mean	Std. Deviation
STEM	Lecture /Discussion /Seminar	Lower Level, General Education	6,525	4.06	0.67
		Lower Level, Specialization	7,731	4.10	0.67
		Upper Level, General Education	1,427	4.09	0.65
		Upper Level, Specialization	6,828	4.16	0.66
		Graduate/Professional	2,120	4.24	0.56
		Total	24,631	4.12	0.66
	Laboratory & Studio	Lower Level, General Education	2,073	4.04	0.64
		Lower Level, Specialization	2,699	4.12	0.64
		Upper Level, General Education	258	3.95	0.66
		Upper Level, Specialization	1,711	4.17	0.66
		Graduate/Professional	139	4.04	0.58
		Total	6,880	4.10	0.64
	Activity /Field /Media /Practicum	Lower Level, General Education	540	4.11	0.61
		Lower Level, Specialization	301	4.14	0.64
		Upper Level, General Education	104	4.18	0.59
		Upper Level, Specialization	386	4.21	0.58
		Graduate/Professional	149	4.04	0.67
		Total	1,480	4.14	0.62
	Total	Lower Level, General Education	9,138	4.06	0.66
		Lower Level, Specialization	10,731	4.10	0.66
		Upper Level, General Education	1,789	4.08	0.65
		Upper Level, Specialization	8,925	4.16	0.65
		Graduate/Professional	2,408	4.21	0.57
		Total	32,991	4.11	0.65
Education	Lecture /Discussion /Seminar	Lower Level, General Education	788	4.36	0.53
		Lower Level, Specialization	1,256	4.39	0.55
		Upper Level, General Education	461	4.30	0.62
		Upper Level, Specialization	3,959	4.30	0.64
		Graduate/Professional	4,678	4.39	0.57
		Total	11,142	4.35	0.59
	Laboratory &Studio	Lower Level, General Education	41	3.73	0.67
		Lower Level, Specialization	38	4.31	0.68
		Upper Level, General Education	11	4.23	0.44
		Upper Level, Specialization	124	4.29	0.66
		Graduate/Professional	52	4.40	0.51
		Total	266	4.23	0.66

	Activity /Field /Media /Practicum	Lower Level, General Education	195	4.44	0.53
		Lower Level, Specialization	303	4.23	0.69
		Upper Level, General Education	70	4.04	0.84
		Upper Level, Specialization	1,250	4.29	0.67
		Graduate/Professional	910	4.32	0.62
		Total	2,728	4.30	0.66
	Total	Lower Level, General Education	1,024	4.35	0.55
		Lower Level, Specialization	1,597	4.36	0.58
		Upper Level, General Education	542	4.26	0.65
		Upper Level, Specialization	5,333	4.30	0.65
		Graduate/Professional	5,640	4.38	0.58
		Total	14,136	4.34	0.61
Business	Lecture /Discussion /Seminar	Lower Level, General Education	965	4.16	0.60
		Lower Level, Specialization	1,851	4.13	0.62
		Upper Level, General Education	1,705	4.16	0.60
		Upper Level, Specialization	5,998	4.20	0.59
		Graduate/Professional	2,969	4.25	0.57
		Total	13,488	4.19	0.59
	Laboratory &Studio	Lower Level, General Education	24	4.25	0.65
		Lower Level, Specialization	33	3.94	0.75
		Upper Level, General Education	7	3.22	0.54
		Upper Level, Specialization	35	4.03	0.79
		Graduate/Professional	20	3.93	0.72
		Total	119	3.99	0.75
	Activity /Field /Media /Practicum	Lower Level, General Education	195	4.29	0.51
		Lower Level, Specialization	274	4.11	0.57
		Upper Level, General Education	112	4.06	0.57
		Upper Level, Specialization	608	4.21	0.59
		Graduate/Professional	281	4.14	0.64
		Total	1,470	4.18	0.59
	Total	Lower Level, General Education	1,184	4.18	0.59
		Lower Level, Specialization	2,158	4.12	0.61
		Upper Level, General Education	1,824	4.15	0.60
		Upper Level, Specialization	6,641	4.20	0.59
		Graduate/Professional	3,270	4.24	0.58
		Total	15,077	4.19	0.59
Liberal Arts	Lecture /Discussion /Seminar	Lower Level, General Education	1,245	4.24	0.57
		Lower Level, Specialization	71	4.31	0.55
		Upper Level, General Education	658	4.31	0.55
		Upper Level, Specialization	116	4.45	0.46

		Graduate/Professional	12	4.39	0.34
		Total	2,102	4.27	0.56
	Laboratory & Studio	Lower Level, General Education	9	3.90	0.63
		Lower Level, Specialization			
		Upper Level, General Education	6	4.36	0.40
		Upper Level, Specialization	4	4.30	0.50
		Graduate/Professional			
		Total	19	4.13	0.56
	Activity /Field /Media /Practicum	Lower Level, General Education	100	4.18	0.58
		Lower Level, Specialization	4	4.44	0.50
		Upper Level, General Education	71	4.41	0.42
		Upper Level, Specialization	55	4.46	0.45
		Graduate/Professional	1	5.00	
		Total	231	4.32	0.52
	Total	Lower Level, General Education	1,354	4.23	0.57
		Lower Level, Specialization	75	4.32	0.54
		Upper Level, General Education	735	4.32	0.54
		Upper Level, Specialization	175	4.45	0.46
		Graduate/Professional	13	4.44	0.37
		Total	2,352	4.28	0.55
Other	Lecture /Discussion /Seminar	Lower Level, General Education	21,709	4.24	0.57
		Lower Level, Specialization	7,117	4.28	0.57
		Upper Level, General Education	4,954	4.29	0.56
		Upper Level, Specialization	17,678	4.32	0.58
		Graduate/Professional	6,745	4.30	0.59
		Total	58,203	4.28	0.57
	Laboratory & Studio	Lower Level, General Education	563	4.28	0.61
		Lower Level, Specialization	1,516	4.34	0.60
		Upper Level, General Education	147	4.33	0.53
		Upper Level, Specialization	1,548	4.33	0.61
		Graduate/Professional	303	4.33	0.57
		Total	4,077	4.33	0.60
	Activity /Field /Media /Practicum	Lower Level, General Education	4,371	4.28	0.56
		Lower Level, Specialization	1,549	4.40	0.55
		Upper Level, General Education	724	4.35	0.54
		Upper Level, Specialization	2,525	4.33	0.59
		Graduate/Professional	901	4.24	0.60
		Total	10,070	4.31	0.57
	Total	Lower Level, General Education	26,643	4.25	0.57
		Lower Level, Specialization	10,182	4.31	0.58

		Upper Level, General Education	5,825	4.30	0.56
		Upper Level, Specialization	21,751	4.32	0.58
		Graduate/Professional	7,949	4.30	0.59
		Total	72,350	4.29	0.58
Total	Lecture /Discussion /Seminar	Lower Level, General Education	31,232	4.20	0.59
		Lower Level, Specialization	18,026	4.19	0.63
		Upper Level, General Education	9,205	4.24	0.59
		Upper Level, Specialization	34,579	4.26	0.61
		Graduate/Professional	16,524	4.31	0.58
		Total	109,566	4.24	0.60
	Laboratory & Studio	Lower Level, General Education	2,710	4.09	0.64
		Lower Level, Specialization	4,286	4.20	0.63
		Upper Level, General Education	429	4.08	0.64
		Upper Level, Specialization	3,422	4.25	0.64
		Graduate/Professional	514	4.24	0.59
		Total	11,361	4.19	0.64
	Activity /Field /Media /Practicum	Lower Level, General Education	5,401	4.27	0.57
		Lower Level, Specialization	2,431	4.31	0.60
		Upper Level, General Education	1,081	4.29	0.58
		Upper Level, Specialization	4,824	4.30	0.61
		Graduate/Professional	2,242	4.25	0.62
		Total	15,979	4.28	0.59
	Total	Lower Level, General Education	39,343	4.20	0.60
		Lower Level, Specialization	24,743	4.21	0.63
		Upper Level, General Education	10,715	4.24	0.59
		Upper Level, Specialization	42,825	4.27	0.61
		Graduate/Professional	19,280	4.30	0.59
		Total	136,906	4.24	0.61

Research Question 3

Dependent Variable: I really wanted to take this course regardless of who taught it (non-doctoral)

Discipline	Course Type	Student Type	N	Mean	Std. Deviation
STEM	Lecture /Discussion /Seminar	Lower Level, General Education	4,305	2.99	0.52
		Lower Level, Specialization	4,948	3.37	0.50
		Upper Level, General Education	1,157	3.02	0.55
		Upper Level, Specialization	2,942	3.45	0.50
		Graduate/Professional	140	3.45	0.45
		Total	13,492	3.24	0.55
	Laboratory & Studio	Lower Level, General Education	739	3.13	0.53
		Lower Level, Specialization	1,374	3.51	0.46
		Upper Level, General Education	130	3.19	0.53
		Upper Level, Specialization	453	3.58	0.49
		Graduate/Professional	5	3.50	0.32
		Total	2,701	3.40	0.52
	Activity /Field /Media /Practicum	Lower Level, General Education	743	3.11	0.59
		Lower Level, Specialization	519	3.46	0.53
		Upper Level, General Education	99	3.27	0.63
		Upper Level, Specialization	257	3.61	0.53
		Graduate/Professional	15	3.65	0.53
		Total	1,633	3.31	0.60
	Total	Lower Level, General Education	5,787	3.03	0.53
		Lower Level, Specialization	6,841	3.41	0.50
		Upper Level, General Education	1,386	3.05	0.56
		Upper Level, Specialization	3,652	3.48	0.50
		Graduate/Professional	160	3.47	0.45
		Total	17,826	3.27	0.55
Education	Lecture /Discussion /Seminar	Lower Level, General Education	375	3.33	0.57
		Lower Level, Specialization	1,122	3.62	0.39
		Upper Level, General Education	152	3.32	0.47
		Upper Level, Specialization	2,165	3.57	0.44
		Graduate/Professional	1,181	3.51	0.43
		Total	4,995	3.54	0.45
	Laboratory & Studio	Lower Level, General Education	18	3.39	0.51
		Lower Level, Specialization	28	3.46	0.46
		Upper Level, General Education	8	3.50	0.49
		Upper Level, Specialization	63	3.38	0.43
		Graduate/Professional	8	3.83	0.36
		Total	125	3.44	0.46

	Activity /Field /Media /Practicum	Lower Level, General Education	116	3.32	0.56
		Lower Level, Specialization	291	3.61	0.48
		Upper Level, General Education	85	3.82	0.59
		Upper Level, Specialization	676	3.66	0.45
		Graduate/Professional	351	3.60	0.45
		Total	1,519	3.62	0.48
	Total	Lower Level, General Education	509	3.33	0.57
		Lower Level, Specialization	1,441	3.61	0.41
		Upper Level, General Education	245	3.50	0.57
		Upper Level, Specialization	2,904	3.59	0.44
		Graduate/Professional	1,540	3.53	0.44
		Total	6,639	3.56	0.46
Business	Lecture /Discussion /Seminar	Lower Level, General Education	873	3.29	0.43
		Lower Level, Specialization	2,026	3.34	0.42
		Upper Level, General Education	829	3.24	0.45
		Upper Level, Specialization	4,818	3.39	0.44
		Graduate/Professional	869	3.52	0.43
		Total	9,415	3.37	0.44
	Laboratory& Studio	Lower Level, General Education	58	2.83	0.48
		Lower Level, Specialization	32	3.33	0.53
		Upper Level, General Education	5	2.93	0.92
		Upper Level, Specialization	28	3.50	0.55
		Graduate/Professional	29	3.67	0.39
		Total	152	3.22	0.60
	Activity /Field /Media /Practicum	Lower Level, General Education	148	3.12	0.47
		Lower Level, Specialization	294	3.24	0.46
		Upper Level, General Education	81	3.28	0.53
		Upper Level, Specialization	536	3.40	0.48
		Graduate/Professional	113	3.60	0.45
		Total	1,172	3.34	0.49
	Total	Lower Level, General Education	1,079	3.24	0.45
		Lower Level, Specialization	2,352	3.32	0.43
		Upper Level, General Education	915	3.24	0.46
		Upper Level, Specialization	5,382	3.39	0.44
		Graduate/Professional	1,011	3.53	0.43
		Total	10,739	3.36	0.45
Liberal Arts	Lecture /Discussion /Seminar	Lower Level, General Education	3,508	2.86	0.55
		Lower Level, Specialization	86	3.24	0.60
		Upper Level, General Education	851	3.01	0.53
		Upper Level, Specialization	68	3.47	0.49

		Graduate/Professional	35	3.32	0.40
		Total	4,548	2.91	0.56
	Laboratory& Studio	Lower Level, General Education	19	3.32	0.55
		Lower Level, Specialization	5	3.65	0.41
		Upper Level, General Education	3	3.03	0.31
		Upper Level, Specialization	3	3.50	0.65
		Graduate/Professional			
		Total	30	3.36	0.53
	Activity /Field /Media /Practicum	Lower Level, General Education	446	2.84	0.54
		Lower Level, Specialization	19	3.40	0.58
		Upper Level, General Education	44	3.01	0.58
		Upper Level, Specialization	13	3.63	0.71
		Graduate/Professional	2	3.32	0.32
		Total	524	2.90	0.57
	Total	Lower Level, General Education	3,973	2.86	0.55
		Lower Level, Specialization	110	3.29	0.60
		Upper Level, General Education	898	3.01	0.53
		Upper Level, Specialization	84	3.50	0.53
		Graduate/Professional	37	3.32	0.39
		Total	5,102	2.91	0.56
Other	Lecture /Discussion /Seminar	Lower Level, General Education	18,237	3.18	0.50
		Lower Level, Specialization	8,437	3.61	0.52
		Upper Level, General Education	4,736	3.23	0.50
		Upper Level, Specialization	12,529	3.52	0.52
		Graduate/Professional	1,848	3.54	0.51
		Total	45,787	3.37	0.54
	Laboratory & Studio	Lower Level, General Education	966	3.58	0.52
		Lower Level, Specialization	2,084	3.76	0.48
		Upper Level, General Education	199	3.66	0.58
		Upper Level, Specialization	1,305	3.78	0.48
		Graduate/Professional	53	3.92	0.56
		Total	4,607	3.73	0.50
	Activity /Field /Media /Practicum	Lower Level, General Education	4,220	3.26	0.58
		Lower Level, Specialization	2,648	3.69	0.52
		Upper Level, General Education	792	3.48	0.63
		Upper Level, Specialization	2,203	3.67	0.55
		Graduate/Professional	317	3.63	0.57
		Total	10,180	3.49	0.60
	Total	Lower Level, General Education	23,423	3.21	0.52
		Lower Level, Specialization	13,169	3.65	0.52

		Upper Level, General Education	5,727	3.28	0.53
		Upper Level, Specialization	16,037	3.56	0.53
		Graduate/Professional	2,218	3.56	0.53
		Total	60,574	3.42	0.56
Total	Lecture /Discussion /Seminar	Lower Level, General Education	27,298	3.12	0.52
		Lower Level, Specialization	16,619	3.50	0.51
		Upper Level, General Education	7,725	3.18	0.51
		Upper Level, Specialization	22,522	3.49	0.50
		Graduate/Professional	4,073	3.52	0.47
		Total	78,237	3.33	0.54
	Laboratory & Studio	Lower Level, General Education	1,800	3.37	0.58
		Lower Level, Specialization	3,523	3.65	0.49
		Upper Level, General Education	345	3.46	0.61
		Upper Level, Specialization	1,852	3.71	0.50
		Graduate/Professional	95	3.81	0.50
		Total	7,615	3.59	0.54
	Activity /Field /Media /Practicum	Lower Level, General Education	5,673	3.20	0.59
		Lower Level, Specialization	3,771	3.61	0.53
		Upper Level, General Education	1,101	3.45	0.64
		Upper Level, Specialization	3,685	3.62	0.53
		Graduate/Professional	798	3.61	0.50
		Total	15,028	3.45	0.59
	Total	Lower Level, General Education	34,771	3.15	0.54
		Lower Level, Specialization	23,913	3.54	0.51
		Upper Level, General Education	9,171	3.22	0.54
		Upper Level, Specialization	28,059	3.52	0.51
		Graduate/Professional	4,966	3.54	0.48
		Total	100,880	3.37	0.55

Research Question 4

Dependent Variable: I really wanted to take this course regardless of who taught it (Doctoral)					
Discipline	Course Type	Student Type	N	Mean	Std. Deviation
STEM	Lecture /Discussion /Seminar	Lower Level, General Education	6,525	3.08	0.47
		Lower Level, Specialization	7,731	3.37	0.48
		Upper Level, General Education	1,427	3.12	0.49
		Upper Level, Specialization	6,828	3.49	0.45
		Graduate/Professional	2,120	3.60	0.45
		Total	24,631	3.33	0.50
	Laboratory & Studio	Lower Level, General Education	2,073	3.16	0.50
		Lower Level, Specialization	2,699	3.45	0.48
		Upper Level, General Education	258	3.18	0.49
		Upper Level, Specialization	1,711	3.57	0.43
		Graduate/Professional	139	3.36	0.59
		Total	6,880	3.38	0.50
	Activity /Field /Media /Practicum	Lower Level, General Education	540	3.19	0.47
		Lower Level, Specialization	301	3.44	0.50
		Upper Level, General Education	104	3.27	0.56
		Upper Level, Specialization	386	3.60	0.50
		Graduate/Professional	149	3.42	0.51
		Total	1,480	3.38	0.52
	Total	Lower Level, General Education	9,138	3.10	0.48
		Lower Level, Specialization	10,731	3.39	0.48
		Upper Level, General Education	1,789	3.14	0.49
		Upper Level, Specialization	8,925	3.51	0.45
		Graduate/Professional	2,408	3.57	0.47
		Total	32,991	3.34	0.50
Education	Lecture /Discussion /Seminar	Lower Level, General Education	788	3.19	0.48
		Lower Level, Specialization	1,256	3.58	0.40
		Upper Level, General Education	461	3.40	0.44
		Upper Level, Specialization	3,959	3.51	0.43
		Graduate/Professional	4,678	3.51	0.46
		Total	11,142	3.49	0.45
	Laboratory & Studio	Lower Level, General Education	41	3.03	0.39
		Lower Level, Specialization	38	3.44	0.47
		Upper Level, General Education	11	3.40	0.36
		Upper Level, Specialization	124	3.39	0.49
		Graduate/Professional	52	3.66	0.43
		Total	266	3.40	0.49

	Activity /Field /Media /Practicum	Lower Level, General Education	195	3.55	0.61
		Lower Level, Specialization	303	3.58	0.50
		Upper Level, General Education	70	3.58	0.49
		Upper Level, Specialization	1,250	3.60	0.43
		Graduate/Professional	910	3.58	0.48
		Total	2,728	3.59	0.47
	Total	Lower Level, General Education	1,024	3.25	0.53
		Lower Level, Specialization	1,597	3.58	0.43
		Upper Level, General Education	542	3.42	0.45
		Upper Level, Specialization	5,333	3.53	0.43
		Graduate/Professional	5,640	3.52	0.47
		Total	14,136	3.51	0.46
Business	Lecture /Discussion /Seminar	Lower Level, General Education	965	3.21	0.45
		Lower Level, Specialization	1,851	3.26	0.43
		Upper Level, General Education	1,705	3.17	0.43
		Upper Level, Specialization	5,998	3.36	0.42
		Graduate/Professional	2,969	3.51	0.41
		Total	13,488	3.35	0.44
	Laboratory& Studio	Lower Level, General Education	24	3.27	0.25
		Lower Level, Specialization	33	3.23	0.39
		Upper Level, General Education	7	3.07	0.50
		Upper Level, Specialization	35	3.51	0.52
		Graduate/Professional	20	3.43	0.39
		Total	119	3.34	0.43
	Activity /Field /Media /Practicum	Lower Level, General Education	195	3.23	0.40
		Lower Level, Specialization	274	3.29	0.43
		Upper Level, General Education	112	3.27	0.46
		Upper Level, Specialization	608	3.35	0.46
		Graduate/Professional	281	3.57	0.44
		Total	1,470	3.36	0.46
	Total	Lower Level, General Education	1,184	3.22	0.44
		Lower Level, Specialization	2,158	3.27	0.43
		Upper Level, General Education	1,824	3.18	0.43
		Upper Level, Specialization	6,641	3.36	0.42
		Graduate/Professional	3,270	3.51	0.41
		Total	15,077	3.35	0.44
Liberal Arts	Lecture /Discussion /Seminar	Lower Level, General Education	1,245	3.08	0.69
		Lower Level, Specialization	71	3.37	0.46
		Upper Level, General Education	658	3.16	0.64
		Upper Level, Specialization	116	3.48	0.49

		Graduate/Professional	12	3.40	0.62
		Total	2,102	3.14	0.67
	Laboratory& Studio	Lower Level, General Education	9	3.33	0.49
		Lower Level, Specialization			
		Upper Level, General Education	6	3.52	1.06
		Upper Level, Specialization	4	3.76	0.55
		Graduate/Professional			
		Total	19	3.48	0.70
	Activity /Field /Media /Practicum	Lower Level, General Education	100	3.07	0.69
		Lower Level, Specialization	4	3.55	0.62
		Upper Level, General Education	71	3.64	0.57
		Upper Level, Specialization	55	3.66	0.44
		Graduate/Professional	1	3.79	
		Total	231	3.40	0.66
	Total	Lower Level, General Education	1,354	3.08	0.69
		Lower Level, Specialization	75	3.38	0.47
		Upper Level, General Education	735	3.21	0.65
		Upper Level, Specialization	175	3.54	0.48
		Graduate/Professional	13	3.43	0.61
		Total	2,352	3.17	0.67
Other	Lecture /Discussion /Seminar	Lower Level, General Education	21,709	3.16	0.48
		Lower Level, Specialization	7,117	3.54	0.50
		Upper Level, General Education	4,954	3.31	0.48
		Upper Level, Specialization	17,678	3.48	0.50
		Graduate/Professional	6,740	3.53	0.52
		Total	58,198	3.36	0.52
	Laboratory & Studio	Lower Level, General Education	563	3.58	0.61
		Lower Level, Specialization	1,516	3.85	0.50
		Upper Level, General Education	147	3.81	0.56
		Upper Level, Specialization	1,546	3.75	0.55
		Graduate/Professional	303	3.97	0.49
		Total	4,075	3.78	0.54
	Activity /Field /Media /Practicum	Lower Level, General Education	4,371	3.22	0.61
		Lower Level, Specialization	1,549	3.71	0.56
		Upper Level, General Education	724	3.52	0.57
		Upper Level, Specialization	2,520	3.66	0.56
		Graduate/Professional	901	3.65	0.54
		Total	10,065	3.47	0.62
	Total	Lower Level, General Education	26,643	3.18	0.51
		Lower Level, Specialization	10,182	3.62	0.52

		Upper Level, General Education	5,825	3.35	0.51
		Upper Level, Specialization	21,744	3.52	0.52
		Graduate/Professional	7,944	3.56	0.53
		Total	72,338	3.40	0.55
Total	Lecture /Discussion /Seminar	Lower Level, General Education	31,232	3.14	0.49
		Lower Level, Specialization	18,026	3.44	0.49
		Upper Level, General Education	9,205	3.25	0.49
		Upper Level, Specialization	34,579	3.46	0.47
		Graduate/Professional	16,519	3.53	0.48
		Total	109,561	3.36	0.51
	Laboratory & Studio	Lower Level, General Education	2,710	3.25	0.55
		Lower Level, Specialization	4,286	3.59	0.52
		Upper Level, General Education	429	3.41	0.60
		Upper Level, Specialization	3,420	3.65	0.50
		Graduate/Professional	514	3.75	0.58
		Total	11,359	3.53	0.55
	Activity /Field /Media /Practicum	Lower Level, General Education	5,401	3.23	0.59
		Lower Level, Specialization	2,431	3.61	0.55
		Upper Level, General Education	1,081	3.48	0.57
		Upper Level, Specialization	4,819	3.60	0.52
		Graduate/Professional	2,242	3.60	0.51
		Total	15,974	3.47	0.58
	Total	Lower Level, General Education	39,343	3.16	0.51
		Lower Level, Specialization	24,743	3.48	0.51
		Upper Level, General Education	10,715	3.28	0.51
		Upper Level, Specialization	42,818	3.49	0.48
		Graduate/Professional	19,275	3.54	0.48
		Total	136,894	3.39	0.52