

THE ANALYSIS OF TEACHING EFFECTIVENESS EVALUATION  
PROGRAMS IN GEOGRAPHY DEPARTMENTS  
IN THE UNITED STATES

DISSERTATION

Presented to the Graduate Council of  
Texas State University-San Marcos  
in Partial Fulfillment  
of the Requirements

for the Degree

Doctor of PHILOSOPHY

by

Brian J. Cooper, M.S.

San Marcos, Texas  
May 2012

UMI Number: 3524584

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent on the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3524584

Copyright 2012 by ProQuest LLC.

All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 - 1346

THE ANALYSIS OF TEACHING EFFECTIVENESS EVALUATION  
PROGRAMS IN GEOGRAPHY DEPARTMENTS  
IN THE UNITED STATES

Committee Members Approved:

---

Dr. Lawrence Estaville, Chair

---

Dr. Brock J. Brown

---

Dr. Richard A. Earl

---

Dr. Emily M. Payne

Approved:

---

Dr. J. Michael Willoughby  
Dean of the Graduate College

**COPYRIGHT**

by

Brian J. Cooper

2012

## **FAIR USE AND AUTHOR'S PERMISSION STATEMENT**

### **Fair Use**

This work is protected by the Copyright Laws of the United States (Public Law 94-553, section 107). Consistent with fair use as defined in the Copyright Laws, brief quotations from this material are allowed with proper acknowledgement. Use of this material for financial gain without the author's express written permission is not allowed.

### **Duplication Permission**

As the copyright holder of this work, I, Brian J. Cooper, refuse permission to copy in excess of the "Fair Use" exemption without my written permission.

## ACKNOWLEDGEMENTS

I would like to sincerely thank the members of my dissertation committee: Chair, Dr. Lawrence E. Estaville, and committee members Dr. Brock J. Brown, Dr. Richard A. Earl, and Dr. Emily M. Payne for their encouragement and guidance in the completion of this dissertation. I would also like to extend my heartfelt thanks to my parents, Jack and Susan Cooper, for their unwavering belief in my abilities and the emotional support they provided during my academic career and life. I would also like to thank my sister, Sara, for inspiring me to achieve my potential. I would also like to thank Dr. Byron Augustin, Dr. Philip Suckling, Angelika Wahl, and Allison Glass-Smith for their guidance and assistance during my professional development. Finally, I thank the fantastic students of Texas State University-San Marcos who were enrolled in my classes during my career here for making this achievement possible.

This manuscript was submitted on 18 March 2012.

## TABLE OF CONTENTS

	<b>Page</b>
ACKNOWLEDGEMENTS.....	v
LIST OF TABLES.....	viii
ABSTRACT.....	x
CHAPTER	
I. INTRODUCTION.....	1
Background of Study.....	2
Significance of Study.....	4
Problem Statement.....	5
II. LITERATURE REVIEW.....	7
The Changing Academic Culture.....	7
Recommended Good Practices in Evaluation of Teaching Effectiveness.....	15
Theoretical Framework.....	24
III. METHOD.....	29
Research Hypotheses.....	30
Study Area.....	31
Timeframe.....	34
Data Collection.....	34
Data Analysis.....	37
Analysis of Variance.....	37
Chi Square Tests.....	38
Qualitative Analysis.....	38
IV. ANALYSIS AND FINDINGS.....	42
Faculty Survey Data.....	42
Analysis of Variance.....	61
Chi Square Analysis.....	63

Chairperson Interview Data .....	66
The Mandate for Evaluation of Teaching Effectiveness.....	68
The Development of the Teaching Evaluation System and Instruments.....	70
The Definition of Quality Teaching.....	72
The Value Placed on Quality Teaching by Department and Institution .....	74
Remedies and Thresholds for Poor Teaching Evaluations .....	76
Differences in Teaching Evaluation Systems between Nontenured and Tenured Faculty.....	80
Teaching as Part of the Total Faculty Workload .....	82
Teaching Performance and Tenure/Promotion Decisions .....	86
The First Qualitative Working Hypothesis .....	88
Discussion of Faculty Survey Responses and Chairperson Interview Responses .....	90
The Second Qualitative Working Hypothesis.....	93
Discussion of Error and Bias .....	95
 V. CONCLUSIONS AND FUTURE RESEARCH.....	97
Conclusions.....	97
Faculty Survey Data Analysis.....	99
Chair Interview Data Analysis.....	100
Faculty Survey and Chair Interview Data Comparison .....	101
Hypotheses Outcomes and Theoretical Contributions.....	101
Future Research .....	104
 APPENDIX	
A. Sample Departments for Faculty Survey Distribution.....	106
B. Teaching Evaluation Survey for Geography Faculty .....	107
C. Interview Questions for Department Chairs/Heads .....	112
 WORKS CITED .....	114



## LIST OF TABLES

Table	Page
1. Salient Characteristics of 13 Sources of Evidence of Teaching Effectiveness.....	21
2. Total Faculty Responses by Census Division and Program Level .....	43
3. Number of Respondents by Faculty Rank by Department Level .....	44
4. Number of Courses Taught in a Typical Semester by Number of Tenure-Track and Tenured Faculty by Department Level .....	44
5. Tenure-Track and Tenured Faculty Responses Regarding Typical Semester Class Sizes .....	45
6. Faculty Perceptions of Value Placed on Quality Teaching by Their Institution by Department Level.....	46
7. Faculty Perceptions of Value Placed on Quality Teaching by Their Department by Department Level.....	47
8. Combined Number of Teaching Evaluation Methods by Department Level .....	49
9. Usage Frequency of Teaching Evaluation Methods by Department Level .....	51
10. Department Development of Teaching Evaluation Methods by Department Level...53	
11. Faculty Responses about Formative Teaching Evaluation Methods by Department Level .....	55
12. Faculty Responses about Summative Teaching Evaluation Methods by Department Level .....	56
13. Faculty Responses regarding Weighting of Teaching Evaluation in Tenure/Promotion Decisions by Department Level .....	57
14. Faculty Responses regarding Weighting of Teaching Evaluation in Merit Decisions by Department Level.....	59
15. Faculty Responses concerning the Existence of Programs in the Department or Institution to Help Improve Teaching Effectiveness .....	60
16. One-Way Analysis of Variance for Institutional Value, Departmental Value, and Total Teaching Evaluation Methods by Nine Census Divisions .....	61

17. Mean Number of Total Evaluation Methods by Census Region .....	62
18. Chi Square Test: Department Type and Perceived Value Placed on Quality Teaching by Institution .....	63
19. Chi Square Test: Department Type and Perceived Value Placed on Quality Teaching by Department.....	64
20. Chi Square Test: Department Type and Total Number of Teaching Evaluation Methods.....	65
21. Chairperson Responses about the Mandate for Teaching Evaluation .....	69
22. Means of Chairperson and Faculty Perceived Value Placed on Quality Teaching by Department and Institution .....	75
23. Frequency of Poor Teaching Remedies by Department Level .....	78
24. Mean Number of Remedies for Poor Teaching by Department Level .....	78
25. Chair Responses by Department Level about Thresholds for Remediating Poor Teaching.....	80
26. Frequency and Type of Differences in Teaching Evaluation Systems for Nontenured and Tenured Faculty by Department Type .....	80
27. Mean, Maximum, and Minimum Percentages for Teaching, Scholarship, and Service by Department Level.....	83
28. Matrix of Agreement between Faculty and Chairs .....	92

## ABSTRACT

# THE ANALYSIS OF TEACHING EFFECTIVENESS EVALUATION PROGRAMS IN GEOGRAPHY DEPARTMENTS IN THE UNITED STATES

by

Brian J. Cooper

Texas State University-San Marcos

May 2012

SUPERVISING PROFESSOR: LAWRENCE E. ESTAVILLE

Teaching evaluation systems are an important part of the total evaluation of faculty in geography departments in the United States. As demands for accountability for teaching effectiveness continue to emerge from many groups, it has become increasingly important for geography departments to develop systems that not only provide teaching effectiveness data to these groups that are aligned with the recommendations for best practices regarding the evaluation of teaching effectiveness from the literature. Using the theoretical framework of the theory of organizational culture, this mixed-method study investigated aspects of the teaching evaluation systems at geography departments in the

United States. This research was particularly focused on any geographical or department-level differences among the departments included in the study.

The quantitative findings of this research showed that geographical differences are minimal, but there were significant differences among geography departments regarding the value placed on quality teaching by institution and department, as well as the total number of evaluation methods. The qualitative analysis revealed that there had been a substantial lack of consultation of the teaching evaluation literature in developing these systems, resulting in considerable variability among the teaching evaluation systems. The qualitative analysis also showed a significant lack of agreement between faculty and chairpersons regarding numerous aspects of the teaching evaluation systems.

## CHAPTER I

### INTRODUCTION

In 2006, the U.S. Department of Education published a report on the state of higher education in the United States. This report examined numerous issues in higher education, one of which was the need for greater accountability in higher education (U.S. Department of Education 2006). The recommendations in this report relating to accountability include:

“creation of a consumer-friendly information database on higher education with useful, reliable information on institutions, coupled with a search engine to enable students, parents, policymakers and others to weigh and rank comparative institutional performance” (U.S. Department of Education 2006, 21).

Some states have used this recommendation to shape policy. In 2009, for instance, the Texas Legislature unanimously passed House Bill 2504. This bill contains numerous requirements for displaying information about tuition costs, syllabi, curriculum vitae of faculty on the Websites of public colleges and universities in that state (Texas 2009). One of the more compelling requirements of this legislation appears in part 3(h) of Sec. 51.974:

Institutions of higher education included in this section shall conduct end-of-course student evaluations of faculty and develop a plan to make evaluations available on the institution’s website (Texas 2009).

This legislation represents a greater change that has been ongoing for decades: the growing demand for accountability of faculty in teaching effectiveness in higher education in the United States.

Seldin (2006c, xiii) wrote in the preface to his book, *Evaluating Faculty Performance*, that, “[f]aculty members are being held accountable, as never before, for how well they do their jobs.” Arreola (2001), Cannon (2001), and Knapper (2001) echoed Seldin’s assertion. Seldin (2006a) also emphasized that teaching, along with research and service, is one of the three most important areas in which faculty performance is evaluated at most colleges and universities in the United States. The evaluation of teaching effectiveness, a topic wrought with controversy, is the specific focus this research within the discipline of geography in higher education in the United States.

### **Background of Study**

Berk (2006) agreed with Seldin (2006a) and concluded that evaluation of teaching effectiveness remains a key component of faculty evaluation systems at most colleges and universities in the United States, although it is not the only or leading way in which faculty are evaluated. The volumes of literature on the evaluation of teaching effectiveness show that numerous disciplines such as business, the natural sciences, sociology, psychology, mathematics, and engineering have had myriad of studies conducted on the ways in which these disciplines evaluate teaching. However, the literature pertaining to the evaluation of teaching effectiveness in higher education

geography programs is noticeably thin. The few articles that address the quality of teaching in higher education geography, such as Chalkley et al. (2000), mentioned the need for an evaluation system but lack in-depth discussions on what such a process should entail and how it should be employed.

Does a need exist for the discipline of geography to specifically examine the way it evaluates the teaching effectiveness of faculty? If so, how are such evaluation data subsequently used in geography departments? The literature about teaching effectiveness evaluation systems provides compelling arguments in favor of both questions. Arreola, in his book, *Developing a Comprehensive Faculty Evaluation System* (2000), conducted a literature review pertaining to the common myths and preconceptions about student evaluations of teaching. Arreola verified the preconception that teachers of math and science courses have a greater difficulty in receiving good ratings via student evaluation and underscored that, “[r]atings tend to be higher for the humanities and social science disciplines as compared to the physical science and engineering disciplines” (2000, 82-83). Most geographers would agree that the discipline of geography spatially analyzes social, cultural, and physical phenomena, with numerous geography departments having faculty who specialize in and teach courses on subjects about one of these broad categories or some combination of them. Given the unique nature of the discipline of geography in this regard, it is imperative to examine the methods of evaluating teaching effectiveness, and how the data gathered by these methods are used for formative (feedback, improvement of teaching) purposes as well as for summative (tenure/promotion/merit) decisions.

## Significance of Study

Seldin (1984) wrote extensively in his book, *Changing Practices in Faculty Evaluation*, about the effects of the difficult economic climate of the early 1980s on higher education, specifically how the poor economy led to a significant increase in the demands for accountability of faculty in terms of job performance from a variety of stakeholders in higher education. Such an atmosphere surrounding higher education in the early 1980s that Seldin explained may be seen again, perhaps in a larger way, during the current fiscal crises since 2008 that have affected the nation, its states, and higher education, in particular.

Once again, the dark clouds of budget cuts by state legislatures loom over colleges and universities across the United States. “Higher education,” Seldin elucidated (1984, 12), “is a labor-intensive industry, with 70 to 80 percent of most budgets going for salaries.” A quarter of a century later, President Graham Spanier of Pennsylvania State University reiterated the same argument in opposing proposed higher education cuts (Danahy 2011). These budget issues will possibly be accompanied by more calls from the stakeholders (e.g., governments, coordinating boards, trustees, donors, students) for accountability for faculty in terms of their job performance. The discipline of geography would be wise to stay ahead of these calls for accountability through an examination of the evaluation systems it uses to assess teaching effectiveness.

The development of a set of best practices for the evaluation of teaching effectiveness in geography departments in colleges and universities across the United States should assist the discipline in preparing for these coming demands. Moreover,



such best practices would also serve the discipline well by being a system that would be thorough, reliable, and valid, based on recommendations found in the literature as well as those corroborated in practice. Of critical importance in these efforts is that they take into account geography's uniqueness as a discipline, as well as meshing well into the evaluation systems already in place in geography departments across the country.

### **Problem Statement**

The purpose of this research is to address the following question: Are the methods and utilization of the evaluation of teaching effectiveness in geography departments at colleges and universities in the United States consistent among the discipline along with the recommendation of good practice of teaching effectiveness evaluation found in the literature for evaluating teaching effectiveness for the discipline? To answer this overarching question, I propose utilizing a mixed—quantitative and qualitative—method. This research should not only uncover how geography departments evaluate their faculty but also some of the possible influences on those decisions. The unique culture within each department may affect the teaching effectiveness evaluation systems, but other factors such as geographic location, due to possible influence from state governments, require analysis. Also, teaching is but one component of the overall evaluation of faculty job performance. While the evaluation of scholarship and service specifically are beyond the scope of this research, departments may use the evaluation data pertaining to teaching effectiveness differently because of programmatic missions and goals, especially in the level of degrees granted. Sizes of

department faculty or types of programs may cause differences in evaluation systems. Varying faculty rank and/or teaching load, possibly showing growing specialization within departments, may affect teaching assessment systems. I seek to analyze how and why geography departments currently evaluate teaching effectiveness to continually promote the ultimate goal of maintaining and improving quality teaching in higher education geography while strengthening it in response to growing demands for accountability.

## CHAPTER II

### LITERATURE REVIEW

The literature clearly shows the interrelationships of the issues of accountability and evaluation of teaching effectiveness. However, to comprehend current practices in place and develop research frameworks about them, it is fundamental to understand the cultures of higher education as a profession, individual institutions, and the discipline of geography. A comprehension of these cultures and the forces that shape them should provide avenues to inform and improve teaching practice and its assessment in geography.

#### **The Changing Academic Culture**

It seems that few discussions of evaluation programs or methods can proceed without addressing the issue of accountability. Arreola (2000, 10-11) affirmed that "...in some measure, the entire faculty evaluation movement has grown out of the larger issue of accountability in education." Accountability denotes "...the requirement to demonstrate responsible actions to external constituencies" (Berdahl, Altbach, and

Gumport 1999, 5). Berdahl and McConnell (1999) drew from Dressel (1980), in a discussion of the nature of accountability to elaborate on this definition, viewing it as a method by which performance is held responsible through a continuous goal-driven process of justifying resource allocation, gathering evidence of good practice, engaging in efforts to improve, and/or efficiently streamline practice.

Institutions of higher education and their faculties are currently being held accountable in unprecedented ways, with calls for indicators of performance coming from a variety of groups within American society (Gaither, Nedwek, and Neal 1994; Moses and Boas 1998; Altbach 1999; Berdahl and McConnell 1999; McGuinness 1999; Tien 1999; Alexander 2000; Knapper 2001; Middaugh 2001; Gates et al. 2002; Seldin 2006c; Estaville 2010). Basic economic issues surrounding higher education have driven this demand for accountability. Seldin (1984) explained how the economic hardships of the early 1980s affected higher education, resulting in calls from groups such as the tax-paying public, alumni and donors, trustees, state legislatures, and students for accountability from higher education and, specifically, about faculty performance.

Gaither, Nedwek, and Neal (1994) wrote about how the public's demands for accountability forced state governments into action, particularly in respect to funding. Institutions of higher education have increasingly found that their annual budgets are the most significant way that state governments can hold them accountable for their decisions (Berdahl and McConnell 1999; Middaugh 2001; Bogue and Hall 2003). Not only are institutions being held accountable, but they "are moving into a period when they will be expected to provide not only data on the attainment of defined outcomes...but also evidence that results have been gained at 'reasonable cost'" (Berdahl and McConnell

1999, 86). Fiscal crises, therefore, increase competition for ever-scarcer funds among institutions of higher education.

McGuinness (1999) postulated that the worse the economic crisis, the greater the number of attempts to enforce accountability by governments. Institutions of higher learning, of course, have to compete for funding with other services, such as K-12 education and Medicare, as outlined by Zusman (1999) in her analysis of the impacts of the economic recession in the early 1990s on higher education budgets. Competition for funds may force changes that may presently seem unpalatable for higher education and its faculty (Tien 1999). Punctuating the outcomes of the reduced budgets in the early 1990s, Zusman (1999, 116) declared that "...the programs cut have been identified as academically weak, high cost, and duplicative and as having low market demand or as being less central to institutional mission or state need." Zusman predicted that the humanities and social sciences stand to experience greater losses from more cost-conscious approaches. During the large higher education budgets cuts in the early 1980s, Aubrecht (1984) championed the need for administrators to incorporate the concept of market value and worth of faculty to institutions, alongside merit, in evaluating faculty performance. Academic programs that can best demonstrate their worth to higher education, to society as a whole, and to those demanding accountability with solid data accentuating efficient, effective utilization of resources to meet mission-driven goals will be better positioned to survive budgetary challenges.

The financial difficulties driving the calls for accountability will have other, far-reaching implications for higher education and its faculty. The very culture of higher education will change as its past autonomy erodes (Alexander 2000). Autonomy and

accountability seem to be located on parallel sliding continua where "...in practice, in cases in which more accountability is required, less autonomy remains" (Berdahl, Altbach, and Gumport 1999, 5). Altbach observed that the continued growth of higher education in both expanse and expense has exacerbated the battle between autonomy and accountability (1999b). The higher education faculty members, thus, resist attempts to undercut their autonomy and change their professional culture through the vehicle of accountability (Altbach 1999b).

Concerns have arisen about the effects of accountability on one of higher education's most sacred concepts: academic freedom. Berdahl, Altbach, and Gumport (1999) defined academic freedom as the fundamental right of faculty to seek knowledge without having to fear for their jobs, no matter where that pursuit may lead them, even if it happens to offend political or social groups of any kind. "Some forms of external control or even subtle efforts to influence teaching, learning, or research may endanger intellectual freedom," concerned Berdahl and McConnell (1999, 71). But the authors continued by drawing a distinction between the concepts of academic freedom and autonomy. They enumerated, for example, program termination, mission statements, and even antidiscrimination regulations in terms of employment practices as actions that, if done properly, would not constitute violations of academic freedom (Berdahl and McConnell 1999).

Academic freedom's strong linkage to tenure, however, brings the issues of performance accountability to the forefront. Indeed, some states have tried to place restrictions on tenure or eliminate it altogether in an effort to be able to more easily terminate faculty seen as performing poorly (O'Neil 1999; Weber 1999). One of the

ways in which these desired changes are manifesting themselves is through increasing use of part-time faculty, nontenured lectureships, and the creation of nontenure track campuses (O’Neil 1999; D’Andrea and Gosling 2005). O’Neil (1999) argued the tenure system could certainly be improved, especially in regards to addressing those who abuse the system. Certainly, though, with or without tenure, the protection of academic freedom is important. D’Andrea and Gosling (2005) posed the concern that academic freedom allows faculty to pursue activities that a more corporate culture would view as “loss leaders,” revealing a clear distinction between an organization of higher education and a traditional corporate organization. However, both Altbach (1999) and Rhodes (1999) saw academic freedom remaining a key component of higher education, as well as one that has support from government and university administration.

The view of students as “customers” or “consumers” epitomizes the shift that threatens higher education’s historical cultural identity (Moses and Boas 1998). Altbach (1999b) argued that the roots of this change came from the rapid growth of higher education enrollments after World War II. Society tasked higher education to educate a workforce that would be better suited for an economy that was expanding in both size and complexity. Weber attributed the more recent changes in the restructuring of American higher education to rising competitiveness rooted in globalization that forces higher education to view students as consumers (1999). Altbach (1999a, 291) declared that, “[s]tudent consumerism is a central part of the ethos of American higher education.” The combination of enrollment and tuition growth has made the costs of higher education a “mass-market” issue (Moses and Boas 1998). Substantial increases in the cost of higher education have driven this consumerism, and stakeholders demand accountability

to verify their money is being spent responsibly (Moses and Boas 1998; Knapper 2001).

The calls for accountability only reinforce this shift toward a more business-like approach in which institutions of higher education are looked upon as “corporations” (D’Andrea and Gosling 2005). Such “corporate mentality” has already become evident at the administration level in many higher education institutions as the higher-level administrators increasingly come from backgrounds other than education (Altbach 1999). These changes will become increasingly important as many institutions of higher education turn to the private sector for funding to close shortfalls resulting from government budget cuts. Accountability will undergird private sector funding and will result in an even greater shift away from the traditional professional culture of higher education to a more corporate model (Altbach 1999b; Zusman 1999; Weber 1999).

A fundamental disconnect between what faculty do and what faculty are *expected* to do has further complicated the issue of accountability. Middaugh (2001) claimed that this dissonance has resulted from higher education faculty communicating poorly what they do to those who wish to hold them accountable. The higher education literature is replete with definitions of faculty productivity that comprise three basic components: teaching, scholarship, and service (Seldin 1984; Birnbaum 1988; Centra 1993; Altbach 1999a; Weber 1999; Arreola 2000; Berk 2006; Seldin 2006a). These three assessment cornerstones of higher education faculty, especially the teaching component, will satisfy the purposes of this research.

Understanding the differences in how these individual components are *valued* by the various groups involved in the issue of accountability, particularly teaching versus scholarship, is imperative. Research institutions have historically considered teaching not



as important as scholarship (Gaither, Nedwek, and Neal 1994; Moses and Boas 1998; Berdahl and McConnell 1999; Zusman 1999; Altbach 1999a; Middaugh 2001; D'Andrea and Gosling 2005). The reasons for this scholarship-teaching relationship are strongly incentive-based in terms of faculty income, rank, and prestige. Gaither, Nedwek, and Neal (1994) summarize five conclusions from an article by Jacobson (1992) in the early 1990s that indicate a positive relationship between research productivity and increases in faculty compensation. Moses and Boas (1998) discussed the investment interjected by colleges and universities to encourage research that result in prestige bestowed upon institutions and faculty scholars. Echoed throughout the higher education literature are the close ties between research productivity and promotion, tenure, and prestige for individual faculty (Zusman 1999; Middaugh 2001; D'Andrea and Gosling 2005).

Weber stated that a consistent complaint of university administrators is a perceived greater allegiance of faculty to their disciplines rather than their institutions (1999). Middaugh (2001) saw this as a manifestation of the emphasis on research and the prestige awarded to faculty who are exceedingly productive in this regard. Moses and Boas (1998) recognized differences in emphasis on research by institution, affirming that research institutions were the first to place a premium value on research, but that over the years this research mission has moved down the hierarchy of institutional type to even some community colleges. Others reiterated these institutional differences in terms of responsibilities and expectations of faculty at each different institutional type (Altbach 1999a; Chalkley, Fournier, and Hill 2000). The type of academic discipline also influences the balance of research versus teaching (Middaugh 2001). Indeed, D'Andrea

and Gosling (2005) concluded that a distinct connection between excellence in teaching and a discipline does not exist.

The literature clearly states the many groups demanding accountability see the focus on research at the expense of teaching as misplaced, ineffective, and inefficient. The focus of the calls for accountability in higher education is therefore on teaching quality (Gaither, Nedwek, and Neal 1994; Altbach 1999a; Zusman 1999; Middaugh 2001; D'Andrea and Gosling 2005). As teaching loads have declined and as many faculty have become more removed from the classroom, dissatisfaction with teaching quality in institutions of higher education has led to increased calls for accountability from the public (Gaither, Nedwek, and Neal 1994; D'Andrea and Gosling 2005; Estaville 2010). Legislators and policymakers have listened to these calls and used the tools at their disposal, primarily funding, to demand systems that will show evidence that institutions of higher education are providing quality classroom instruction by evaluating performance and productivity (Berdahl and McConnell 1999; Zusman 1999; Bogue and Hall 2003). States have long been major leaders in educational reform at all levels; the power their governments wield in terms of legislation and the ability to demand accountability from entities such as state coordinating boards cannot be overstated (McGuinness 1999). Faculty productivity in quality instruction will continue to raise pertinent questions, especially those that are tied to the stated and assumed missions of colleges and universities. However, a central question remains: how should institutions of higher education evaluate an activity as complex as teaching?

### **Recommended Good Practices in Evaluation of Teaching Effectiveness**

Despite the extensive discussion about the demands for accountability, accountability should not be the only concern when developing a system to evaluate teaching effectiveness. Bogue and Hall (2003) argued that, no matter what is being evaluated, assuring quality is a moral activity. D'Andrea and Gosling (2005, 66) elaborated: “[t]he ultimate aim of professional development activity is to improve the quality of teaching and learning to the benefit of the students.” “Assessing faculty performance in teaching,” underlined Higgerson (2006, 35), “yields information that is essential to enhancing student learning, faculty development, program quality, and institutional success.” Seldin (2006c) reiterated the benefits to these outcomes when the evaluation process is done well, adding that a thoughtful, understandable approach to overall faculty performance evaluation allows decisions such as promotion, tenure, and other employment-related decisions to be accomplished more fairly and rationally. The underlying benefits of faculty assessment thus go well beyond just placating those who seek more accountability from higher education.

The feasibility of evaluating teaching effectiveness is a topic of controversy in the literature. Some question whether teaching assessment can be done at all, arguing that quality of teaching is tied to learning assessment, a continuing conundrum, to say the least (Weber 1999). The increased focus on learning will certainly shape approaches to evaluation of teaching effectiveness, but as more data are gathered, these data can be used to improve evaluation instrument and program design (Knapper 2001; Berk 2006; Higgerson 2006). Given the frequency of students transferring between institutions (taking general education courses at one institution, then taking major courses at another),

Gaither, Nedwek, and Neal (1994) argued that judging an entire institution based on what individual students have learned is unfair. This position is intriguing but does not offer practical solutions. Some outright claimed it is impossible to evaluate teaching because the education profession has yet to provide a definition of quality teaching (Altbach 1999a). While there seems to be no *universal* definition of quality teaching, the many definitions that exist share numerous characteristics (Centra 1993; Arreola 2000; Chalkley, Fournier, and Hill 2000; Berk 2006; Seldin 2006a; Estaville 2010). Yet, as they developed assessments for teaching effectiveness, institutions incorporated good teaching evaluation practices found in the literature, including many of these definitions and characteristics. Those who dismiss the feasibility of evaluating teaching effectiveness in the literature, nevertheless, will not deter those demanding accountability from higher education, nor will they allow higher education faculty to develop professionally in a manner that improves student learning.

Before implementing any form of evaluation regarding faculty performance, understanding and tackling faculty resistance is crucial. Arreola (2000) identified faculty resistance, along with administrator apathy, a related issue, as one of the two major reasons why faculty evaluation programs fail. The problems with outlining a clear definition of teaching are factors driving resistance (Fite 2006). The introduction of evaluation systems has often been met with considerable resistance as the culture of academia has historically been one with the individual faculty member working in an environment free from scrutiny (Gaither, Nedwek, and Neal 1994). These authors also maintained that faculty have resisted these changes as unwanted intrusions of corporate mentality into academic culture. A fear of the process or a lack of trust among faculty

and between faculty and the administration also causes faculty resistance to the process (Arreola 2000). The paucity of formal teaching training that faculty and evaluators have received may also trigger this resistance to pedagogical assessment (Fite 2006).

Minimizing faculty resistance is key to the process of evaluating teaching effectiveness (Gates et al. 2002; Higgerson 2006). Arreola (2000) listed guidelines for minimizing errors and resistance in the faculty evaluation process. This list includes such suggested remedies as establishing an office of faculty development and advisory board independent of a higher administration office (including possibly using a consultant independent of the faculty and administration); developing programs that are tailored specifically to help faculty improve every aspect of teaching effectiveness evaluated by the system; using several different sources to comprehensively provide evaluation data, ensuring validity of the system and instruments; making certain the evaluation data are controlled properly to reinforce trust between faculty and administration and creating a reward system to positively encourage improvement and/or excellence, specifically by tying the evaluation of teaching effectiveness as closely as possible to the tenure/promotion/merit decisions in the department or institution. Fite (2006) stressed clear expectations, faculty involvement throughout the process of evaluation, education and training for both faculty and administrators in regard to the evaluation process, and a strong connection between the evaluation process and mentoring and development programs that enable faculty to grow professionally as teachers. Seldin (2006a) cautioned that, despite the care that may be taken, evaluation programs must also be designed to account for unusual fluctuations in evaluation data across time and departments, a type of “grade inflation” of evaluation where evaluators reward below-

average efforts with above-average scores, and the dangers of an over-reliance on numbers and subsequent treatment of those numbers as objective facts.

If care is not taken to address faculty resistance and faculty concerns or if there are flaws in the system, the consequences could be serious. All those involved in the process should also be concerned with the legal implications of any faculty evaluation system. As Berdahl and McConnell (1999, 80) warned: “[f]aculty members may sue over dismissal, appointment, tenure, and accessibility to personnel records.” Grave issues arise if flawed evidence from poor systems guides decisions about retention, promotion, tenure, and merit (Seldin 1984). “In 1990 the Supreme Court ruled that university tenure reviews are not shielded by any special privilege from the general laws of evidence” (Fite 2006, 185). Certainly no institution of higher education, or department, for that matter, wishes to have an entire system used to make such important career decisions declared publicly flawed. Keeping faculty involved in the entire process while working to minimize the resistance helps reduce the likelihood of having to fight such battles in court.

The discussion of faculty resistance touched upon the use of teaching effectiveness evaluation data in two different ways: faculty professional development as teachers and faculty career advancement in the profession. These two types of evaluation or evaluation decisions are, respectively: *formative evaluation decisions* and *summative evaluation decisions*. Knapper (2001) outlined that summative evaluation more accurately addresses the accountability demands, while formative evaluation is more reflective for purposes of informing and improving teaching practice. Benenson (1994) added a temporal aspect to these different types of evaluation, elaborating that the

formative approach takes place during a program or term and the summative approach takes place after a program or term has been completed. Centra (1993, 5-6) presented an additional perspective to formative evaluation, suggesting that “informative feedback may be a better term.” Centra also emphasized that such evaluation or feedback should take the form of constructive criticism offering solutions for improvement. Both types, he stated, will contribute to addressing the calls for accountability. Knapper (2001, 6) stated that “there is considerable overlap between these two types of evaluation.” Zakrajsek (2006) emphasized the importance of formative evaluation, with the recommendation that summative data be used to provide guidance and direction for future formative evaluation. Berk (2006, 11) stated “[t]he various sources of evidence for teaching effectiveness may be employed for either formative or summative decisions or both.” Institutions of higher education must determine what kind of decisions will be made as a result of any evaluation program due to these intended uses determining the methods of evaluation that are most appropriate.

The selection of evaluation methods not only depends on the types of decisions the data gathered by those methods will be used to make, but also by the appropriateness of evaluators. Are students equipped with the skills to evaluate teaching effectiveness? Should colleagues evaluate each other? Do administrators have the ability to evaluate a venture in which few of them have training? Can faculty be expected to evaluate themselves objectively? Can other groups, such as recent graduates, alumni ideally utilizing the education bestowed upon them in the workforce, the employers of those alumni provide useful feedback about teaching effectiveness? Should the whole process be left to neutral, external observers who might further strain already-tightened budgets?

Seldin (2006a) explained the first four groups (students, colleagues, administrators, and self-evaluators) can provide critical perspectives to the evaluation of teaching effectiveness, provided the appropriate questions are asked, and the appropriate aspects of teaching each group is qualified to evaluate are selected. Centra (1993) discussed these same four groups at length in his development of a program for evaluating teaching effectiveness. While the specific benefits and deficiencies of each group as evaluators of teaching effectiveness are beyond the scope of this study, the literature lucidly stressed multiple sources should be used (Centra 1993; Gaither, Newdek, and Neal 1994; Berk 2006; Fite 2006; Seldin 2006b). These authors justify this strategy by indicating that each evaluating group is not qualified to assess all aspects of teaching effectiveness; if the intention is to evaluate the overall effectiveness of teaching, multiple sources become necessary. Arreola (2000) devised a matrix showing which aspects of a four-pronged definition of effective teaching students, peers, and the department chairperson are qualified to evaluate. He underscored none of the three groups was qualified to make accurate evaluations of more than two of the four aspects of the definition of teaching being used. Halonen and Ellenberg (2006) summarized typical errors made by students, faculty, and administrators in the many steps of the evaluation process. These authors made a recommendation to create an evaluation program that brings multiple sources of data together to avoid these errors. Berk (2006) offered a list of 13 sources that can produce evaluation of teaching effectiveness data for formative, summative, and programmatic decisions. Reproduced in Table 1, this framework also lists the evidence providers, the type of measurement instrument, and the evaluation decision makers appropriate according to the evidence provided.



**Table 1.** Salient Characteristics of 13 Sources of Evidence of Teaching Effectiveness.

<i>Source of Evidence</i>	<i>Type(s) of Measure</i>	<i>Who Provides Evidence</i>	<i>Who Uses Evidence</i>	<i>Type(s) of Decision*</i>
Student Ratings	Rating scale	Students	Instructor/administrator/ curric. committee	F/S/P
Peer Ratings	Rating scale	Peers	Instructor/administrator	F/S
External Expert Ratings	Rating scale	Outside experts	Instructor	F/S
Self-Ratings	Rating scale	Instructor	Instructor/administrator	F/S
Videos	Rating scale	Instructor/peers	Instructor/peers	F/S
Student Interviews	Rating scale	Students	Instructor/administrator	F/S
Exit/Alumni Ratings	Rating scale	Graduates	Instructor/curric. committee	F/P
Employer Ratings	Rating scale	Graduates' employers	Administrator/curric. committee	P
Administrator Ratings	Rating scale	Administrator	Administrator/ promotions committee	S
Teaching Scholarship	Judgmental review	Instructor	Administrator	S
Teaching Awards	Judgmental review	Instructor	Faculty committee/ administrator	S
Learning Outcome Measures	Tests, projects, simulations	Students	Instructor	F
Teaching Portfolio	Most of the above	Instructor/students/ peers	Promotions committee	S

\*F = formative S = summative P = program

Reproduced from Berk, Ronald A. 2006. *Thirteen strategies to measure college teaching*. Sterling, VA: Stylus Publishing, LLC, 15.

Table 1 brings together many of the themes that run throughout the literature in regard to good teaching effectiveness evaluation practice. The methods chosen must match appropriately with the instrument(s) being used, the provider(s) of the evidence, the user(s) of the evidence, and the purpose of that use.

Validity and reliability of the evaluation of teaching are essential topics of discussion throughout the literature. Arreola (2000, 80-81) dispelled some of the myths surrounding student evaluation forms by indicating that many of the forms in use are invalid and unreliable because they do not follow “the rigorous psychometric and

statistical procedures required to produce a professional, well-developed student rating form.” Pallett (2006) related an anecdote about individuals calling the IDEA Center, a firm that develops student evaluation instruments for use by institutions of higher education. He discovered from these conversations that locally developed instruments frequently have not been tested for validity and reliability. Pallett reiterated the need for psychometric development, a lack of which can breed distrust in the evaluation instrument and system. Those systems that have been developed according to these guidelines are valid and reliable. Validity, in particular, has been reinforced when comparing these ratings with ratings from other sources of evidence such as peer, external, and self-ratings (Arreola 2000). Arreola described reliability as having both a temporal and a source of evidence component. Important questions include: 1) is there stability of response among evaluation scores between different times in a given course; and 2) what sort of consistency exists among the responses? Both stability and consistency must be demonstrated in order for an instrument to be classified as reliable. Validity and reliability are both important because, while reliability implies consistency of measurement, validity shows the instrument is appropriately measuring teaching effectiveness as intended (Berk 2006).

Logical validity and empirical validity are distinct from each other, as well. Logical validity, “requires judgment on the content validity of the instrument,” while empirical validity, “require[s] the use of criterion measures against which...ratings may be compared” (Arreola 2000, 96). Validity depends on constructing an instrument that is specifically tailored to measure what it intends to measure, using evaluators best suited to provide that evidence, and subsequently measuring what it was intended to with a high

degree of correlation between the criteria and the ratings (Arreola 2000). The evaluation literature is consistent that validity and reliability are essential to any quality program evaluating teaching effectiveness. Seldin (2006b) summarized that one of the errors that most commonly leads to unreliability and invalidity of evaluation programs is lack of consistency and quality in the standards being used in the evaluation process.

Gates et al. (2002, 6) defined evaluation as “the step in the assessment process in which measures of quality and productivity are examined against some standard of performance.” Centra (1993, 6) declared that “standards define the desired level of achievement.” But these standards must be specifically tailored to the evaluation program (Gaither, Newdek, and Neal 1994). Often they are linked to institutional standards that are found in union contracts or faculty handbooks (Arreola 2000). Centra (1993) advocated tying the standards directly to the definition of teaching being used. Arreola (2000) outlined methods that the statements of standards and achievement level can be converted to numerical data for quantitative analysis. Berk (2006, 205) showed methods by “[s]ubscale and total scale scores and/or global item scores can provide a meaningful picture of the students’, the department’s, and the instructor’s ratings of teaching performance.” While global items can often be controversial, experts approve of them because they “are correlated moderately with student learning” and “provide an easily understandable summary of teaching” (Berk 2006, 205). Using these rankings in a norm-referenced approach that ranks faculty can provide performance ratings for a particular semester or term (Berk 2006). Pallett (2006) advised using broader benchmarks of performance; often, particularly with student rating scores, evaluators make major decisions due to very minor differences in ranking.

But all groups involved in the process of evaluating teaching effectiveness must be informed of the standards, as well as the criteria for meeting each standard of performance (Seldin 2006b). Deficiencies here can undermine the entire evaluation program. Seldin (2006b) indicated the danger of unclear criteria and standards leaving faculty lost in terms of the performance expected of them. Centra (1993, 7) cautioned that “[r]elative judgments can damage collegiality and cooperation, especially when the comparison group is small and faculty members are competing against each other for promotions and raises.” Higher education can become vulnerable to the legal implications discussed earlier by proceeding haphazardly in this process, with possible failure to clearly inform and improve good teaching practice.

### **Theoretical Framework**

Birnbaum, in his 1988 book *How Colleges Work*, examined institutional culture from the perspective of fictional Huxley College to provide an understanding of how colleges and universities function from an organizational culture perspective.

Universities are unique as organizations, particularly when compared with typical corporate structures and as a manifestation termed “academic culture” (Birnbaum 1988; Gaither, Newdek, and Neal 1994; Altbach 1999b; D’Andrea and Gosling 2005).

“Organizational culture is a powerful way of looking at how people in institutions create social reality through their interactions and interpretations,” according to Birnbaum (1988, 72). I discuss the theory of organizational culture in a broad sense before applying it directly to institutions of higher education.

My research fits directly into Schein's theory of organizational culture. Schein (2006) outlined three levels of culture: artifacts, beliefs and values, and assumptions. Artifacts include, "the visible products of the group...its language; its technology and products; its artistic creations...its published lists of values; its observable rituals and ceremonies" (Schein 2004, 25-26). These artifacts are manifestations of the espoused values and beliefs of the organization. The values represent, "[s]trategies, goals, philosophies (espoused justifications)" (Schein 2004, 26). These values reflect in an organization the right and wrong ways in which work should be done. Members of an organizational culture create, transform, and diffuse these principles. He defined the basic assumptions as, "[u]nconscious, taken-for-granted beliefs, perceptions, thoughts, and feelings... (ultimate source of values and action)" (Schein 2004, 26). These assumptions provide the meaning, and, in turn, are shaped by the values they create and the artifacts used to express these values and assumptions. The levels of culture are constantly interacting with each other to create the culture of the organization.

Birnbaum (1988) identified academic culture in the United States as one in which faculty and administrators work together to provide the public service of accessible high-quality education in an atmosphere of autonomy and academic freedom to contribute to broader understanding and knowledge and to produce graduates who are well-prepared to contribute to the workforce. However, the real professional prestige for faculty in academic culture, argued Gaither, Newdek, and Neal (1994), often comes from research productivity, not teaching; this predilection has manifested itself in the different values placed upon teaching versus research.

An organizational culture may contain subcultures. Hatch (2006, 176) defined *subculture* as, “a subset of an organization’s members that identify themselves as a distinct group within the organization and routinely take action on the basis of their unique collective understandings.” Subcultures can develop based on common interests or as a result of frequent interaction with individuals that, in the end, become a part of the same subculture (Hatch 2006). A hierarchy of cultures can therefore develop within the organization. Gates et al. (2002) discussed the need to understand the heterogeneity resulting from variables such as geographic location, size of institution, and the relationship between those institutions and stakeholders in higher education when choosing how to assess any type of institutional performance. Altbach (1999a) reaffirmed the differences among institutional type in terms of shaping professional responsibilities and expectations. The type of institution, whether it is a community college, liberal arts college, or research university, can result in cultural differentiation among institutions in the United States, according to Birnbaum (1988). In Middaugh’s (2001) mind, differences in institutions of higher education can manifest in the degree of emphasis placed on teaching. Subcultural qualities based on the value placed on teaching by an institution or department may vary from the recommendations of good practices in the evaluation of teaching effectiveness literature.

Birnbaum (1998) also noted that academic disciplines are subcultures within the larger organizational culture of higher education. “University leaders,” according to Weber (1999, 10), “complain regularly that faculty are more faithful to their discipline than to their university,” again underscoring common interests as one of the driving forces behind the formation of subcultures. Middaugh (2001) lamented about these

disciplinary subcultures, which he felt were caused by the rewards of increasing specialization. Middaugh also pointed out that emphasis on teaching can vary according to discipline as well. Thus, examination of the evaluation of teaching effectiveness *within the discipline of geography* is essential because subcultural differences may exist that are unique to the discipline and that may explain variations from recommended good practices of teaching effectiveness evaluation.

Finally, because of subcultures produced by geographic location, dissimilarities may exist among groups of institutions. Birnbaum (1988) explored the intrusive nature of the environment on organizations and explained that outside forces influence (and can be influenced by) those organizations. The environment, for the purposes of this study, will be the state in which geography departments are located. Despite the influence of state governments as forces demanding accountability, there are “great variations among states in size, culture, policies, and structure,” according to McGuinness (1999). The subculture of states, he continued, is often reflected through the institutions of higher education within them. He outlined two broad categories of state characteristics and how these characteristics manifest themselves by influencing colleges and universities within those states. McGuinness contextualized these examples by citing the political science literature on the subcultures of individual state politics (1999). While he acknowledged that states have long been leaders in educational reform and that states often act in ways that imply consensus of direction among them, states often have subtle differences that impact the degree to which they choose to act individually regarding higher education policies. State governments, in response to public outcry, have been one group demanding accountability from institutions of higher education and are using fiscal tools

at their disposal to enforce such demands. State-level analyses and comparisons are therefore critical in understanding the evaluation of teaching effectiveness programs of departments of geography.



## CHAPTER III

### METHOD

My fundamental research question was: Are the methods and utilization of the evaluation of teaching effectiveness in geography departments at colleges and universities in the United States consistent throughout the discipline to be combined with the recommendations of good practice of teaching effectiveness evaluation found in the literature for evaluating teaching effectiveness for the discipline? Models have been developed that fit the recommendations of the literature, and certainly other disciplines have examined their own evaluation of teaching effectiveness programs, but there has been no comprehensive study of what geography departments in the United States are specifically undertaking to evaluate teaching effectiveness. This mixed-method study investigated the evaluation of teaching effectiveness methodology and utilization to discover the nature of those programs and analyze any differences among them to determine the feasibility of developing a model for the entire discipline of geography. In approaching the analysis of teaching effectiveness programs, the four quantitative working hypotheses of this study focused on (1) the associations of important variables noted in the literature that may influence departmental characteristics and (2) the geographic location by state. The qualitative working hypothesis examined the degree to which geography departments are following the recommendations of good practice in the

teaching evaluation literature, thus providing a more comprehensive picture of evaluation of teaching effectiveness programs in the United States.

### **Research Hypotheses**

Eight working hypotheses (six quantitative and two qualitative) provided the basic framework and direction for this research. All five quantitative hypotheses are presented as null hypotheses.

#### Quantitative

(1) No significant geographical differences according to the U.S. Census Divisions exist in the value placed on quality teaching by geography departments in the United States.

(2) No significant geographical differences according to the U.S. Census Divisions exist in the value placed on quality teaching by geography departments' institutions in the United States.

(3) No significant geographical differences according to the U.S. Census Divisions exist in regard to the total number of evaluation methods.

(4) No significant differences according to type of department based on the highest degree offered exist in the value placed on quality teaching by geography departments' institutions in the United States.

(5) No significant differences according to type of department based on the highest degree offered exist in the value placed on quality teaching by geography departments in the United States.

(6) No significant differences according to type of department based on the highest degree offered exist in the total number of teaching evaluation methods utilized by geography departments in the United States.

### Qualitative

(1) The teaching effectiveness evaluation methodology and utilization programs in geography departments in the United States have been developed without significant consultation of the literature pertaining to evaluation of teaching effectiveness, thus resulting in considerable variability throughout the discipline in regard to the application of the principles.

(2) Agreement exists among faculty and department chairs/heads about the characteristics of the evaluation of teaching effectiveness methodology, utilization, and the value placed upon teaching by geography departments in United States and the value placed upon teaching by the institutions in which those departments are located.

### **Study Area**

My study area consisted of the fifty states of the United States and the District of Columbia. Because of the limited number of departments that may be sampled in any

one state, geographic differences were analyzed by state as well as employing statistical analysis on data from the U.S. Census Bureau's nine divisions. Time constraints did not allow for examination of every geography department in the United States. A sampling method therefore selected the sample departments to provide representation of every state and the District of Columbia, along with equal representation of departments classified by highest degree offered within each census division.

The method drew the sample geography departments from the Association of American Geographers' *AAG Guide to Geography Programs in the Americas 2009-2010*, which provides information about departments that consider geography a significant part of their academic identities. The guide also contains information about number of faculty and the highest degree offered by each department. To become a part of the sample, departments had to offer at least a bachelor's degree. The hierarchical procedure I used to construct this department sample was:

- (1) All fifty states and the District of Columbia must be represented by at least one geography department.
- (2) In census divisions with less than twelve states (twelve is the largest number of states in any division), multiple geography departments must come from at least one state.

These multiple departments were selected in a hierarchical manner:

(A) The state or the District of Columbia must have multiple geography departments available for selection.

(B) States with multiple geography departments were ranked according to 2010 Census population data. States with larger populations then had a greater chance

of having more departments offering a larger diversity of degrees at the undergraduate, master's, and doctoral levels. No additional department was allowed to replace a department that is the only available department from a state.

(C) If multiple departments must come from a state, priority was given to those having a different degree level from the first selected department. If there was no other option, multiple departments offering the same degree level was selected.

(3) U.S. Census divisions were the first-order sampling frame. If departments from the states in the divisions did not fill the division sample size, then I drew from departments from states in the same U.S. Census region to complete the sample size.

This procedure was used to create a three (based on highest degree offered: bachelor's programs, master's programs, and doctorate programs) by nine (each census division) matrix of geography departments in the United States, with each of the twenty-seven cells in the matrix filled with four geography departments, thus creating a total sample of 108 geography departments (Appendix A).

The sample for the qualitative portion of this study included 27 department chairs/heads from the departments listed in each of the 27 cells of the matrix (Appendix A). Although every effort was made to provide as wide a geographic distribution as possible, the willingness of the chairs/heads determined which departments are included from each cell.

### **Timeframe**

The study period for my research was the 2010-11 academic year. The analysis did not consider comparing data through time.

### **Data Collection**

I developed two instruments to gather data about teaching effectiveness evaluation programs. The first survey was distributed via e-mail to every faculty member currently listed on each sampled department's Website. In the event of hybrid departments (such as Geography and Anthropology or Geography and Geology), only those faculty indicating themselves as geography faculty received the survey. The sample size for this survey was 1,255 faculty members. With a response rate of 25 percent, a total of 316 of the 1255 faculty members provided responses usable for the quantitative analysis of this study.

As shown in Appendix B, the faculty survey questions included items for quantitative and qualitative analysis:

- (1) Faculty member rank.
- (2) Number of classes taught in a typical semester.
- (3) General description of the size(s) of classes typically taught.
- (4) Faculty perceptions of the value placed on quality teaching by the department.
- (5) Faculty perceptions of the value placed on quality teaching by the institution.

- (6) The type of data gathered by the evaluation instruments (quantitative, qualitative, or both).
- (7) Whether or not faculty are given scores for their teaching effectiveness.
- (8) Whether or not faculty are ranked according to scores.
- (9) Methods used to evaluate teaching effectiveness for the responding faculty member by the department.
- (10) Which of the methods used were developed by the department?
- (11) Were any of the methods developed using the recommendations found in the literature pertaining to the evaluation of teaching effectiveness?
- (12) Which of the methods are intended to provide formative evaluation data?
- (13) Which of the methods are intended to provide summative evaluation data?
- (14) Do defined thresholds that faculty must meet to satisfy the department requirements for effective teaching exist?
- (15) The weight given to teaching effectiveness in the department's promotion/tenure process.
- (16) The weight given to teaching effectiveness in the department's merit process.
- (17) Does the department have formal methods or programs for helping faculty improve low teaching evaluation scores?

(18) Does the institution have formal methods or programs for helping faculty improve low teaching evaluation scores?

The second instrument was questions administered via structured telephone interviews to 27 geography department chairs/heads. As displayed in Appendix C, the interview questions were constructed to provide both qualitative and quantitative data and include similar topics explored by the faculty survey. After each interview was completed, I transcribed the conversation and e-mailed it to the interviewee to verify the accuracy of the responses and to reach goals of the qualitative analysis. I also took notes during each interview that provided a transcription of my immediate interpretations of the responses as they were being given, an essential part of the qualitative data collection process.

While the faculty surveys intended to provide the point of view of those being evaluated in teaching effectiveness evaluation process, the chair/head interviews furnished the viewpoints of the evaluators from their administrative perspectives. The interviews also uncovered many more insights of the philosophies and processes of geography departments' teaching effectiveness evaluation by allowing for follow-up questions. These key informant interviews therefore became a narrative to enable a richer explanation of the teaching effectiveness evaluation process in the selected geography departments.



## **Data Analysis**

The quantitative analysis of this research utilized data gathered from both collection methods—faculty and chairs/heads. This analysis attempted to either prove or disprove the five quantitative working hypotheses. I aggregated the quantitative data from the surveys and interviews into categories based on U.S. state (or District of Columbia), U.S. Census division and department type according to highest degree offered. I derived aggregated data for total number of evaluation methods from the faculty surveys and chair interviews.

### Analysis of Variance

Analysis of variance (ANOVA) is useful when the independent variables are categorical in nature. A one-way ANOVA was run to address the first three quantitative working hypotheses. The dependent variables for the one-way ANOVA were: (1) faculty responses about the perceived value placed on quality teaching by the department, (2) faculty responses about the perceived value placed on quality teaching by the institution, and (3) the total number of methods used to evaluate teaching effectiveness as indicated by the faculty. The data for the first two dependent variables related to the perception of the value placed on quality teaching were gathered using Likert-scale questions with fixed, evenly spaced anchors. Although the nature of these data is ordinal, because the anchors are spaced equally to portray equal intervals between them and the underlying variable being measured is quantitative, I was able to convert these responses to quasi-interval data for ANOVA (Kachigan 1991). Each anchor was assigned a numerical value: 0 for highly unvalued, 1 for unvalued, 2 for average, 3 for valued, and 4

for highly valued. To determine if any geographical differences exist among the data, the dependent variables were analyzed in reference to the independent variable, geographic location according to U.S. Census division.

### Chi Square Tests

Some of the faculty survey data was analyzed using chi-square tests to address the remaining three quantitative working hypotheses. The purpose of a chi-square test is to determine if the frequency distribution of a set of values in a sample is consistent with the expected values. The chi-square test for independence is particularly useful when dealing with categorical variables such as department type.

Chi square tests were performed on the faculty responses to the questions about perceived value placed on quality teaching by institution, perceived value placed on quality teaching by department, and the total number of teaching evaluation methods using department type (bachelor's, master's, and doctoral departments) as the test variable. These tests helped determine if any significant differences exist among the data according to the three department types by generating  $\chi^2$  values,  $\phi$  values (correlations), and p-values. The p-values were compared to a significance level or  $\alpha$  of .05 with higher values being out of the range of accepting the null hypotheses.

### Qualitative Analysis

The analysis of the chair/department head interviews was accomplished using the grounded theory methodology explained by Creswell (1998). The purpose of the grounded theory approach is to generate a theory that explains how “individuals interact, take actions, or engage in a process in response to a phenomenon” (Creswell 1998, 56).

One of the purposes of this study was to examine why geography departments evaluate teaching effectiveness the way that they do in response to the demand for data and accountability from the variety of sources. The grounded theory methodology was thus best suited for this research.

As stated earlier, the purpose of this portion of the qualitative analysis was to provide greater insight into the nature of the evaluation of teaching effectiveness process in the selected departments from an administration point of view. The key here was not necessarily to ascertain how geography departments were conducting teaching evaluations, but, instead, to discover the justifications for why teaching effectiveness was being assessed. This analysis involved the open coding and axial coding methods prescribed by Creswell (1998). Grounded theory studies evolve iteratively as they take place, as data collection and analysis proceed in a *zig-zag* fashion, with collection followed by analysis, then a return to the field for more collection to be followed by further analysis, and so forth. The intent here was to create categories of data with dimensionalized properties, a process known as open coding. Once the categories were considered saturated, a point that was reached after I interviewed and analyzed the responses from the 27<sup>th</sup> chairperson, I employed axial coding to construct a narrative explaining why teaching effectiveness evaluation programs were constructed and utilized in geography departments.

Although it is difficult to specify exactly what this research was looking for in these categories of information, the grounded theory approach required that I disregard any theoretical notions or assumptions I may have in order to let the narrative explain the phenomena as they exist. I also had to set aside any biases or assumptions about how and

why teaching effectiveness evaluation programs were developed and utilized when collecting and analyzing the data. Due to my research on the issues of accountability and the understanding of good practices as recommended by the literature, I could not let these issues influence the questions I pose to the key informants, either in the constructed interview framework or in any follow-up questions I may have posed during any individual interview to gather more information or in the creation of categories of information.

I then compared the interview responses of the department chairs/heads to the faculty responses from the same 27 departments. This appraisal was used to examine the degree of agreement about the components of teaching effectiveness evaluation programs between the faculty and the administration. One of the aspects I carefully examined in this portion of the qualitative analysis was the perception of the value placed upon quality teaching by the department and the institution. The limited sample of department chair/heads made robust quantitative analysis inappropriate. Any large difference between faculty and department chairs/heads regarding these variables on an individual basis was scrutinized closely against recommendations of best practice from the literature to provide possible explanations for the discrepancy.

The overall purpose of the qualitative analysis was to address the qualitative working hypotheses, as well as to either reinforce the conclusions of the quantitative analysis or provide possible explanations of any anomalies that may emerge from the quantitative analysis. The underlying influences revealed by the qualitative analysis informed this research about the possible reasons for the differences between the

recommendations of good practice in the literature and the actual practices of the evaluation of teaching effectiveness in these departments.

## CHAPTER IV

### ANALYSIS AND FINDINGS

#### Faculty Survey Data

With a response rate of 25 percent, a total of 316 of the 1255 faculty members provided responses usable for the quantitative analysis of this study. These 316 responses came from faculty at 101 of the 108 departments sampled in 49 of the 50 states and the District of Columbia; the only state not represented was South Dakota. Every census division is exemplified by at least one response in at least two departments for every program level—doctoral, master’s, and bachelor’s (Table 2). The 101 responding departments, moreover, were divided almost equally among the three department levels: 33 bachelor’s, 34 master’s, and 34 doctoral. A master’s-granting department in the East North Central Census Division had the highest number of faculty responding with nine completed surveys; only one faculty member responded from each of 19 departments. Doctoral departments had a higher mean response total than the master’s departments; the master’s had a higher mean response total than the bachelor’s departments. This relationship is most likely a result of doctoral and master’s departments having more faculty members than bachelor’s departments.

**Table 2.** Total Faculty Responses by Census Division and Program Level.

Census Division	Doctoral Departments (Total Faculty Responses)	Master's Departments (Total Faculty Responses)	Bachelor's Departments (Total Faculty Responses)	Total
New England	4 (13)	4 (13)	4 (5)	12 (31)
Middle Atlantic	4 (12)	4 (17)	4 (5)	12 (34)
East North Central	4 (15)	3 (16)	4 (13)	11 (44)
West North Central	3 (11)	3 (11)	3 (5)	9 (27)
South Atlantic	5 (21)	4 (11)	4 (13)	13 (45)
East South Central	3 (12)	4 (14)	4 (10)	11 (36)
West South Central	4 (17)	3 (10)	4 (9)	11 (36)
Mountain	3 (13)	5 (15)	3 (10)	11 (38)
Pacific	4 (12)	4 (8)	3 (5)	11 (25)
Total	34 (126)	34 (115)	33 (75)	101 (316)

Table 3 shows 316 faculty respondents for each rank by department level. Of the responding faculty members, 277 (88%) held ranks that would typically require a workload that includes teaching, scholarship, and service. Although doctoral departments had the largest number of professors, associate professors, and assistant professors (113) responding in comparison with master's departments (97), and bachelor's departments (67), the overall distribution among the faculty ranks was similar: 31.6% professors, 28.4% associate professors, and 27.5% assistant professors.

Every responding tenure-track and tenured faculty member indicated teaching at least one course per semester (Table 4). Over half of the respondents (152) taught two courses in a typical semester with doctoral departments having the largest number of faculty (92) teaching such a two-course load. Of the 12 faculty members responding they taught five or more courses per semester, none taught at a doctoral-granting department; half held the rank of professor, although five of those professors taught at bachelor's-granting departments.

**Table 3.** Number of Respondents by Faculty Rank by Department Level.

Rank	Bachelor's Department Respondents	Master's Department Respondents	Doctoral Department Respondents	Total Number of Respondents	Percentage of Total Responses
Professor	26	28	47	100	31.6
Associate Professor	24	35	30	90	28.4
Assistant Professor	17	34	36	87	27.5
Lecturer/ Instructor	6	13	11	30	9.4
Adjunct	2	5	2	9	2.8
Total	75	115	126	316	100.0

**Table 4.** Number of Courses Taught in a Typical Semester by Number of Tenure-Track and Tenured Faculty by Department Level (n = 277).

Classes Taught in a Typical Semester	Bachelor's Department Faculty	Master's Department Faculty	Doctoral Department Faculty	Total Number of Faculty
One	1	4	15	20
Two	19	41	92	152
Three	26	26	5	57
Four	16	20	1	37
Five or More	5	7	0	12

A total of 14 of the 20 faculty who responded they teach only one course per semester held the rank of professor; 15 out of these 20 faculty taught at doctoral-granting departments. Table 4 clearly displays, therefore, that the typical teaching load for faculty at doctoral departments is lower than for faculty at master's and bachelor's departments.

I defined class size using three categories: small (30 or less students), medium (31-100 students), and large (more than 100 students). The largest number of faculty,



116 (41.8%), indicated they typically taught medium and small classes in a semester (Table 5). Only four faculty members (1.4%) reported they taught all large classes in a typical semester.

**Table 5.** Tenure-Track and Tenured Faculty Responses Regarding Typical Semester Class Sizes (n = 277).

Class Sizes	Bachelor's Department Responses	Master's Department Responses	Doctoral Department Responses	Total Number of Faculty Responses
All Large Classes	1	1	2	4
All Medium Classes	8	11	3	22
All Small Classes	9	17	9	35
Large and Medium Classes	5	15	9	29
Large and Small Classes	7	10	16	33
Medium and Small Classes	32	35	49	116
Large, Medium, and Small Classes	5	9	24	39

The faculty survey questions regarding the perceived value placed on quality teaching by the institution and by the department were critically necessary for the chi square analysis. Any survey that did not include a response to either of these questions was discarded from the overall analysis. As a result, five surveys were excluded from the overall analysis (321 total raw responses; 316 usable responses). Both value questions used the same Likert scale with five ratings in equal steps. Each rating was converted to

a numerical value for the quantitative analyses: highly unvalued to 1, unvalued to 2, neutral to 3, valued to 4, and highly valued to 5. The mean perceived value placed on quality teaching by the faculty members' institutions was a 3.7 out of 5.0 (Table 6).

**Table 6.** Faculty Perceptions of Value Placed on Quality Teaching by Their Institution by Department Level.

Value	Bachelor's Department Responses (%)	Master's Department Responses (%)	Doctoral Department Responses (%)	Number of Faculty Responses (%)
1 – Highly Unvalued	4 (5.3)	5 (4.3)	9 (7.1)	18 (5.7)
2 – Unvalued	1 (1.3)	13 (11.3)	15 (11.9)	29 (9.2)
3 – Neutral	6 (8.0)	16 (13.9)	24 (19.0)	46 (14.6)
4 – Valued	44 (58.6)	54 (46.9)	63 (50.0)	161 (50.9)
5 – Highly Valued	20 (26.7)	27 (23.5)	15 (11.9)	62 (19.6)
Total	75	115	126	316

A total of 223 faculty members (70.6%) felt their institutions either valued or highly valued quality teaching, as opposed to 47 faculty members (14.9%) who assigned negative ratings of unvalued or highly unvalued. Upon examining the data by department level, the negative values increase and the positive values decrease as a percentage of the total when moving from bachelor's to master's to doctoral departments. The difference between the responses from bachelor's departments and the responses from doctoral departments is especially notable. The faculties at bachelor's departments perceive their institutions value quality teaching more frequently than faculty at master's or doctoral departments.

The faculty responses regarding perceived value placed on quality teaching by the departments yield similar results (Table 7). The average perceived value placed on

quality teaching by the faculty members' departments was slightly higher at 3.92 out of 5.0.

**Table 7.** Faculty Perceptions of Value Placed on Quality Teaching by Their Department by Department Level.

Value	Bachelor's Department Responses (%)	Master's Department Responses (%)	Doctoral Department Responses (%)	Number of Faculty Responses (%)
1 – Highly Unvalued	4 (5.3)	4 (3.5)	11 (8.7)	19 (6.0)
2 – Unvalued	0 (0.0)	7 (6.1)	9 (7.1)	16 (5.1)
3 – Neutral	4 (5.3)	15 (13.0)	20 (15.9)	39 (12.3)
4 – Valued	32 (42.7)	49 (42.6)	58 (46.0)	139 (44.0)
5 – Highly Valued	35 (46.7)	40 (34.8)	28 (22.2)	103 (32.6)
Total	75	115	126	316

Table 7 indicates perceived department values are skewed higher for bachelor's departments than for either master's or doctoral departments. Also, although the number of respondents from doctoral departments was slightly larger than the number of master's respondents and considerably larger than the number of bachelor's respondents, doctoral departments had the lowest number of respondents indicating that quality teaching was highly valued by their departments. Chi-square analysis will examine the statistical significance of these differences, but the trends noticeable in this table bear out the view that teaching has greater value at bachelor's departments and less value at doctoral departments.

When compared to the perceived institution values, a higher number of faculty, 242 respondents (76.6%), felt their departments either valued or highly valued teaching, as opposed to a lower number, 35 respondents (11.1%), who assigned the negative ratings

of unvalued or highly unvalued. These differences may be the result of greater familiarity and frequency of the recognition of quality teaching at the departmental level rather than the institutional level, possibly illustrating the department organizational culture is stronger than the organizational culture of the institution.

Of the 316 respondents, 314 faculty members responded to the question regarding whether or not faculty are rated or scored in the teaching evaluation process. A total of 238 faculty members (75.8%) indicated they were scored in the process; however, 38 (12.1%) specified they were not scored, and the remaining 38 (12.1%) reported they did not know if they were scored in the process. Even if the faculty who did not know were actually scored in the process of teaching evaluation, those scores were most likely not shared with them in any meaningful way. Scores or not, far fewer faculty members responded they were ranked alongside their colleagues as part of the teaching evaluation process. A total of 316 faculty members responded to the ranking question, with only 66 (20.9%) responding that ranking was undertaken. Over 57% (182 faculty) responded “no,” with the remaining 68 responses (21.5%) being “not certain/I don’t know,” once again showing, if any ranking was performed, it was not shared with these faculty.

Five separate survey questions covered the utilization of specific evaluation methods, the development of the methods, and the purpose of the data collected by the methods. The first of these questions asked faculty to indicate which evaluation methods were used to evaluate their teaching effectiveness. Faculty members were provided with a list of 12 teaching effectiveness methods adapted from Berk’s (2006) list in Table 1: student evaluations, peer evaluations, chair evaluations, dean evaluations, external expert evaluations, teaching portfolios, formal self-evaluations, videos, student interviews,

exit/alumni evaluations, learning outcomes measures, and teaching awards. Employer ratings were excluded because it was only to be used for program decisions, not formative or summative faculty assessments. Teaching scholarship was also excluded from the list because of possible confusion it might create; I operated under the assumption that any such scholarship would be evaluated as part of a faculty member's scholarship workload and evaluation. Table 8 displays the number of faculty responses for each combined number of teaching evaluation methods employed by department level. A total of 316 faculty members responded to this question with at least one method being used to evaluate teaching effectiveness. Responding faculty members' teaching effectiveness was evaluated with an average of 3.17 different methods.

**Table 8.** Combined Number of Teaching Evaluation Methods by Department Level.

Number of Methods	Bachelor's Department Responses	Master's Department Responses	Doctoral Department Responses	Number of Faculty Responses
1 method	8	14	27	49
2 methods	18	28	33	79
3 methods	20	25	33	78
4 methods	6	26	15	47
5 methods	10	14	8	32
6-9 methods	13	8	10	31

A total of 267 faculty members (84.5%) responded their teaching effectiveness was evaluated using more than one method. It is not surprising that the number of faculty evaluated by an increasing number of methods decreases; time and expense are certainly major factors influencing how many methods can be used on every faculty member. Important to note is that the total number of methods incorporated to evaluate each faculty member was not consistent within departments; of departments with multiple

faculty responses, only seven (two bachelor's departments, two master's departments, and three doctoral departments) had all responding faculty agreeing on the total number of evaluation methods used. All seven departments also had only two faculty respondents. Out of those seven, only four departments (one bachelor's department and three doctoral departments) exhibited consistency among responding faculty as to which specific methods were included. None of the departments in agreement had more than two responding faculty members.

Faculty members at bachelor's departments indicated greater frequency of use of multiple teaching evaluation methods than at master's or doctoral departments. The best teaching evaluation practices in the literature recommend multiple methods be utilized to evaluate teaching effectiveness. Also, the greater number of methods may enhance the perceptions by faculty members that quality teaching is more highly valued. Chi square analysis will examine the statistical significance of these differences.

The faculty surveys showed each one of the 12 specific teaching effectiveness evaluation methods were used to evaluate at least one faculty member in the sample departments. The frequency of each method's use varied considerably. Table 9 presents the frequency of usage of each of the 12 specific teaching evaluation methods as shown by the faculty responses.

Student evaluations were cited by every responding faculty member as a method in which their teaching effectiveness was evaluated. This finding is not surprising from an accountability standpoint or from an investigation of the voluminous literature about student evaluations. Moreover, all 49 faculty members who reported only one teaching

evaluation method were assessed using student evaluations. Peer evaluation was the only other method applied to appraise more than half of the responding faculty. Teaching awards and chair evaluations followed with 41% and 40%, respectively. Five of the remaining eight methods—dean evaluations, teaching portfolios, formal self-evaluations, exit/alumni evaluations, learning outcomes measures—were each employed less than 20% of the time, and departments rarely used the three other methods—external expert evaluations, videos, and student interviews.

**Table 9.** Usage Frequency of Teaching Evaluation Methods by Department Level.

Evaluation Method	Bachelor's Department Responses	Master's Department Responses	Doctoral Department Responses	Total Number of Faculty Indicating Usage	Percent Usage
Student Evaluations	75	115	126	316	100
Peer Evaluations	39	63	71	173	55
Chair Evaluations	35	59	33	127	40
Dean Evaluations	13	15	4	32	10
External Expert Evaluations	2	3	2	7	2
Teaching Portfolios	19	22	20	61	19
Formal Self-Evaluations	21	24	7	52	16
Videos	1	1	2	4	1
Student Interviews	4	2	5	11	3
Exit/Alumni Evaluations	13	9	10	32	10
Learning Outcomes Measures	16	24	18	58	18
Teaching Awards	30	37	62	129	41

Table 10 presents the responses about particular teaching assessment methods and instruments departments developed by department level. More than two-thirds of the faculty surveyed reported the teaching evaluation methods used by their departments were developed in the departments. More than half of the faculty responded, however, that only three other methods were developed by their departments: exit/alumni evaluations, learning outcome measures, and peer evaluations. Learning outcomes and department alumni are very specific to each discipline or department, so it is no surprise that these methods would be departmentally developed. The substantial percentage of peer evaluation instruments developed by departments may suggest a higher degree of control that faculty in these departments wish to have over how they are evaluated by their colleagues or this aspect of departmental culture.

The faculty members were also asked if the literature about teaching evaluation was consulted in developing the assessment methods utilized. Of the 314 respondents, only 48 faculty persons (15.3%) were able to confirm the methods employed by their departments had been developed according to the recommendations in the teaching evaluation literature. A similar number of faculty, 53 or 16.9%, responded no literature had been consulted. The large majority of the faculty, 213 or 67.8%, admitted they did not know if any literature had been consulted in the development of their department teaching evaluation methods. This outcome highlights a substantial lack of knowledge regarding this particular aspect of the teaching evaluation systems, meaning that faculty may be unaware that potentially flawed instruments or methods are being used to evaluate their teaching.



**Table 10.** Department Development of Teaching Evaluation Methods by Department Level.

Evaluation Method	Bachelor's Department Responses	Master's Department Responses	Doctoral Department Responses	Number of Faculty Indicating Department Developed Method	Percentage of Total Faculty Indicating Method Usage
Student Evaluations	24	18	25	67	21.2
Peer Evaluations	27	27	41	95	54.6
Chair Evaluations	14	22	17	53	41.4
Dean Evaluations	0	0	1	1	3.1
External Expert Evaluations	0	0	0	0	0
Teaching Portfolios	6	4	5	15	24.6
Formal Self-Evaluations	8	6	2	16	30.8
Videos	0	0	0	0	0
Student Interviews	2	1	2	5	45.5
Exit/Alumni Evaluations	9	8	5	22	68.8
Learning Outcomes Measures	11	16	10	37	63.8
Teaching Awards	2	2	4	8	6.2
I Don't Know	2	3	8	13	4.4
None	22	41	39	102	32.6

Table 11 shows the number of faculty indicating which teaching evaluation methods utilized by their departments was included to make formative decisions by department level. Only two methods, videos and external expert evaluation, were incorporated for formative decisions by more than 50 percent of the faculty who stated

the teaching evaluation method was used in their departments. Over 40 percent of the faculty reported no methods were involved in making formative decisions. Berk (2006) emphasized two methods, exit/alumni ratings and learning outcome measures, should be used for only formative purposes; however, the percentages of responding faculty designating those methods were used for any formative decisions are very low. Likewise, according to Berk (2006), the dean evaluations and teaching awards should be incorporated for only summative decisions, yet a small percentage of faculty specified each of these methods were used for formative decisions.

Table 12 exhibits the number of faculty indicating which teaching evaluation methods utilized by their departments were employed to make summative decisions. No method had a 100 percent summative decision usage rate, even those methods (dean evaluations and teaching awards) that should be exercised solely for summative purposes. Departments used five methods (student evaluations, peer evaluations, chair evaluations, formal self-evaluations, student interviews) to make summative decisions about the teaching effectiveness for more than 50 percent of the faculty being evaluated by those methods. Two methods Berk (2006) underscored should be applied for only formative decisions (exit/alumni ratings and learning outcome measures) were used to make summative decisions for substantial numbers of faculty that were evaluated employing these methods.

**Table 11.** Faculty Responses about Formative Teaching Evaluation Methods by Department Level.

Evaluation Method	Bachelor's Department Responses	Master's Department Responses	Doctoral Department Responses	Number of Faculty Indicating Method Usage for Formative Decisions	Percentage of Total Faculty Indicating Method Usage
Student Evaluations	18	18	31	67	21.2
Peer Evaluations	13	25	36	74	42.5
Chair Evaluations	9	13	8	30	23.4
Dean Evaluations	1	1	0	2	6.3
External Expert Evaluations	1	2	2	5	71.4
Teaching Portfolios	2	2	2	6	9.8
Formal Self-Evaluations	5	4	1	10	19.2
Videos	1	1	1	3	75.0
Student Interviews	2	1	1	4	36.4
Exit/Alumni Evaluations	1	0	0	1	3.1
Learning Outcomes Measures	4	9	2	15	25.9
Teaching Awards	0	0	3	3	2.3
I Don't Know	2	7	11	20	6.3
None	41	45	44	130	41.1

**Table 12.** Faculty Responses about Summative Teaching Evaluation Methods by Department Level.

Evaluation Method	Bachelor's Department Responses	Master's Department Responses	Doctoral Department Responses	Number of Faculty Indicating Method Usage for Summative Decisions	Percentage of Total Faculty Indicating Method Usage
Student Evaluations	63	93	108	264	83.5
Peer Evaluations	26	39	35	100	57.5
Chair Evaluations	20	34	18	72	56.3
Dean Evaluations	6	7	2	15	46.9
External Expert Evaluations	1	0	1	2	28.6
Teaching Portfolios	8	9	4	21	34.4
Formal Self-Evaluations	14	14	2	30	57.7
Videos	1	0	0	1	33.3
Student Interviews	3	4	2	9	81.8
Exit/Alumni Evaluations	4	4	2	10	31.3
Learning Outcomes Measures	7	13	8	28	48.3
Teaching Awards	7	10	10	27	20.9
I Don't Know	3	7	6	16	5.1
None	6	6	5	17	5.4

Several trends emerged while examining the responses about teaching evaluation methods employed for summative decisions. Only two methods were identified as being used by departments more often for formative decisions rather than summative decisions: external expert evaluations and videos, although these methods were included very rarely

overall. The remaining methods were selected far more often for summative decisions. Also, far more faculty indicated methods were used for summative decisions rather than formative decisions. These trends suggest summative evaluation is the more dominant form of teaching effectiveness evaluation data in departments.

Faculty members were asked whether the teaching evaluation systems in their departments had defined thresholds that faculty must meet to satisfy department requirements for teaching effectiveness. Only 90 faculty persons (28.6%) responded their departments had such defined thresholds. Nearly a majority of respondents, 156 or 49.5%, specified no thresholds exist, with another 69 faculty (21.9%) reporting they did not know if thresholds exist. These data emphasize expectations in terms of level of performance remained unclear for 70% of responding faculty, certainly an unhealthy situation for faculty evaluation systems and the students taught.

Two faculty survey questions addressed the weighting given to teaching performance regarding to tenure/promotion and merit decisions (Table 13).

**Table 13.** Faculty Responses regarding Weighting of Teaching Evaluation in Tenure/Promotion Decisions by Department Level (n = 311).

Weighting	Number of Bachelor's Department Responses	Number of Master's Department Responses	Number of Doctoral Department Responses	Number of Faculty Responses	Percentage of Total Faculty Responses
0%	0	1	1	2	0.6
1-20%	3	14	32	49	15.8
21-40%	22	42	53	117	37.6
41-60%	18	15	9	42	13.5
61-80%	20	12	3	35	11.3
81-100%	6	2	0	8	2.6
I Don't Know	6	28	29	58	18.6

Tenure and promotion were linked together in one question, as these decisions are analogous to each other; merit raises, on the other hand, can be distributed independently of any tenure or promotion decisions. While 58 respondents did not know if there was a formal weighting for teaching evaluation in terms of the overall evaluation of faculty for tenure and promotion decisions, the total responses in percentages had an almost bell-curve distribution to their answers. The mode of the responses was that teaching evaluation accounted for 21-40 percent of the total evaluation for tenure and promotion decisions. Of the two faculty members who responded teaching evaluation accounted for zero percent of the total tenure/promotion evaluation process, one held the rank of professor at a doctoral department and the other was an associate professor at a master's department. Of the eight faculty persons responded teaching accounted for 81-100 percent of the total tenure/promotion evaluation process, six faculty members of various ranks taught at bachelor's departments, with the remaining two teaching at master's level departments.

More noteworthy trends become evident when the individual department levels are examined. The higher percentage ranges were much more pronounced at the bachelor's departments, and the lower ranges much more punctuated at the doctoral departments. Another trend of note is the substantially fewer "I don't know" responses from the bachelor's departments as compared to the master's and doctoral departments. Not only do bachelor's departments more frequently give amplified weighting to teaching effectiveness in terms of tenure/promotion decisions, but the faculty at these departments seem to be more aware of the greater weight placed on teaching effectiveness in the tenure/promotion process.

The next survey question addressed the weighting of teaching evaluations in the merit decision process (Table 14). The trends observed in the weighting of teaching evaluation in the merit decision process when examined by department level are similar to those in the tenure/promotion process: the higher percentages are typically more acute at the bachelor's departments, and the lower percentages become more represented as the levels increase to master's and doctoral departments. In terms of the specific percentage ranges, the 21-40 percent range once again represented the most frequent selection.

**Table 14.** Faculty Responses regarding Weighting of Teaching Evaluation in Merit Decisions by Department Level (n = 303).

Weighting	Number of Bachelor's Department Responses	Number of Master's Department Responses	Number of Doctoral Department Responses	Number of Faculty Responses	Percentage of Total Faculty Responses
0%	6	7	13	26	8.6
1-20%	2	21	33	56	18.5
21-40%	16	20	37	73	24.1
41-60%	15	14	12	41	13.5
61-80%	10	8	2	20	6.6
81-100%	3	2	0	5	1.7
I Don't Know	23	42	30	82	27.1

A larger number and percentage of respondents did not know any specific weight range for teaching evaluation in the merit decision-making process than in the tenure/promotion process. In fact, the "I don't know" response was the mode of the merit decision question. The lack of awareness by such a considerable number of faculty members raises some questions about whether the faculty have been informed about how merit decisions are made. The responses from these two questions helped to reinforce the view that greater emphasis is placed on teaching at bachelor's departments as opposed to

master's and doctoral departments, with doctoral departments placing the lowest emphasis on teaching effectiveness.

The final two questions of the faculty survey inquired about formal programs provided by departments or institutions that can help faculty improve teaching effectiveness (Table 15).

**Table 15.** Faculty Responses concerning the Existence of Programs in the Department or Institution to Help Improve Teaching Effectiveness.

Department	Response	Does your department have programs to improve teaching effectiveness?	Does your institution have programs to improve teaching effectiveness?
Bachelor's Department Responses	Yes	7	34
	No	53	29
	Not Certain/Don't Know	15	12
Master's Department Responses	Yes	11	57
	No	76	29
	Not Certain/Don't Know	29	29
Doctoral Department Responses	Yes	21	74
	No	83	21
	Not Certain/Don't Know	22	31
Totals	Yes	39	165
	No	212	79
	Not Certain/Don't Know	66	72

Several trends seem unusual at first glance. The higher number of doctoral departments with departmental programs for teaching improvement may be explained by the larger average department size of doctoral departments, which may afford them the resources to provide more formal programs. The higher number of bachelor's departments with no formal programs at either the department or institution level seems odd as well; however, this situation may be explained by smaller average department size



limiting resources for departmental programs and possibly a smaller number of formal programs overall at the institutions in which these departments are located due to any number of reasons, particularly resources available. Overall, these responses show a majority of departments leave the responsibility of formal teaching training or improvement to programs developed by their institutions.

### Analysis of Variance

I performed a one-way analysis of variance (ANOVA) to address the working hypotheses concerning any geographical differences among the 316 faculty responses regarding perceived value placed on quality teaching by their institutions, perceived value placed on quality teaching by their departments, and the total number of evaluation methods used to evaluate teaching effectiveness according to the nine census divisions. The degrees of freedom for the among-groups variance was 8, 307 for the within-groups variance, and 315 for the total variance. Table 16 provides the *F*-ratio results of the one-way analysis of variance, along with the confidence level of each ratio.

**Table 16.** One-Way Analysis of Variance for Institutional Value, Departmental Value, and Total Teaching Evaluation Methods by Nine Census Divisions.

Dependent Variable	<i>F</i> -ratio	Confidence Level
Perceived Value Placed on Quality Teaching by Institution	1.011	.428
Perceived Value Placed on Quality Teaching by Department	1.945	.053
Total Teaching Evaluation Methods	4.920	.000

*F*-ratios were significant at the 95% confidence level. The *F*-ratio for the perceived value placed on quality teaching by institution versus census region was not significant at a level of confidence of 95%, indicating no significant relationship between these two variables. The *F*-ratio for the perceived value placed on quality teaching by department versus census region was almost significant at 95% confidence level, but still resulting in no significant relationship between these two variables. Based on these analyses, I accepted the null hypotheses that no significant geographical differences according to the U.S. Census divisions exist in the value placed on quality teaching by geography departments or their institutions in the United States.

The *F*-ratio for the total number of evaluation methods versus census regions was critical at a level of confidence greater than 99%, indicating a very strong, significant relationship between these two variables. After examining the raw data, I realized one census division, the Mountain division, was most likely responsible for this result due to the much higher average number of total evaluation methods provided by faculty responses from departments in the division, suggesting a possible sample bias (Table 17).

**Table 17.** Mean Number of Total Evaluation Methods by Census Region.

New England	Middle Atlantic	East North Central	West North Central	South Atlantic	East South Central	West South Central	Mountain	Pacific
2.58	3.26	3.30	2.74	3.13	2.83	2.75	4.61	3.00

Based on this exception, I rejected the null hypothesis that no significant geographical differences according to the U.S. Census divisions exist in regard to the total number of evaluation methods.

### Chi Square Analysis

I performed chi square tests to determine if the faculty responses for the perceived value placed on quality teaching by institution, the perceived value placed on quality teaching by department, and the total number of teaching evaluation methods were consistent among the three department types. The chi square tests were performed on responses from tenure-track and tenured faculty only. Lecturers, instructors, and adjunct faculty were omitted from this analysis because of their minimal knowledge about their departments. Table 18 displays the  $\chi^2$  values,  $\phi$  values, and p-values for the chi square tests performed on the bachelor's, master's, and doctoral department faculty responses to the question pertaining to perceived value placed on quality teaching by institution.

The chi square tests between the perceived values placed upon quality teaching by institution for all three department levels provided strong evidence to reject the null hypothesis. Because of these conclusive results, I rejected the null hypothesis that no significant differences according to type of department based on the highest degree offered exist in the value placed on quality teaching by geography departments' institutions in the United States.

**Table 18.** Chi Square Test: Department Type and Perceived Value Placed on Quality Teaching by Institution.

Department Level Comparison	$\chi^2$ values	$\phi$ values	p-values
Bachelor's, Master's, and Doctoral Departments	21.03	.268	.01
Bachelor's and Master's Departments	9.25	.231	.06
Bachelor's and Doctoral Departments	17.73	.303	.00
Master's and Doctoral Departments	8.01	.194	.09

I compared each department level to the others individually to determine if any specific department level was driving these differences. The bachelor's-master's test yielded suggestive evidence to reject the null hypothesis. The master's-doctoral test also yielded suggestive, yet weaker, evidence to reject the null hypothesis. However, the bachelor's-doctoral test provided the strongest evidence to reject the null hypothesis. The responses from bachelor's departments displayed that the value placed on quality teaching by the institution was higher than the responses from the other levels.

Table 19 displays the  $\chi^2$  values,  $\phi$  values, and p-values for the chi square tests performed on the bachelor's, master's, and doctoral department faculty responses to the question pertaining to perceived value placed on quality teaching by department.

**Table 19.** Chi Square Test: Department Type and Perceived Value Placed on Quality Teaching by Department.

Department Level Comparison	$\chi^2$ values	$\phi$ values	p-values
Bachelor's, Master's, and Doctoral Departments	20.24	.261	.01
Bachelor's and Master's Departments	8.35	.220	.08
Bachelor's and Doctoral Departments	19.41	.312	.00
Master's and Doctoral Departments	5.16	.155	.27

The results of these tests were similar to the tests on the perceived value placed on quality teaching by institution. The chi square tests between the perceived values placed on quality teaching by department for all three department levels provided strong evidence to reject the null hypothesis. Because of these decisive outcomes, I rejected the null hypothesis that no significant differences according to type of department based on the highest degree offered exist in the value placed on quality teaching by geography departments in the United States.

I compared each department level to the others individually to determine if any specific department level was causing these variances. The bachelor's-master's test yielded suggestive evidence to reject the null hypothesis. However, the master's-doctoral test yielded no evidence to reject the null hypothesis. The bachelor's-doctoral test once again provided the strongest evidence to reject the null hypothesis. The responses from bachelor's departments underscored the value placed on quality teaching by these department was higher than the responses from the other levels. The value of other professional activities, such as scholarship, may diminish the value of teaching at master's and doctoral departments.

Table 20 displays the  $\chi^2$  values,  $\phi$  values, and p-values for the chi square tests performed on the bachelor's, master's, and doctoral department faculty responses to the total number of teaching evaluation methods. The chi square test between the total number of teaching evaluation methods for all three department levels provided strong evidence to reject the null hypothesis. I thus rejected the null hypothesis that no significant differences according to type of department based on the highest degree offered exist in the total number of teaching evaluation methods utilized by geography departments in the United States.

**Table 20.** Chi Square Test: Department Type and Total Number of Teaching Evaluation Methods.

Department Level Comparison	$\chi^2$ values	$\phi$ values	p-values
Bachelor's, Master's, and Doctoral Departments	24.96	.288	.01
Bachelor's and Master's Departments	13.36	.274	.02
Bachelor's and Doctoral Departments	11.73	.247	.04
Master's and Doctoral Departments	10.49	.218	.06

I once again undertook an internal pair-wise analysis. The bachelor's-master's test and the bachelor's-doctoral test yielded moderate evidence to reject the null hypothesis. The master's-doctoral test produced weaker evidence to reject the null hypothesis. The responses from bachelor's departments displayed that the number of total teaching evaluation methods was higher than the responses from the other levels.

When these results were examined within the context of the results from the tests on responses to the perceived value placed upon quality teaching questions, the impact of the higher number of teaching evaluation methods at bachelor's departments became clearer. The higher perceived values at bachelor's departments made sense because the teaching effectiveness was typically evaluated using more methods. The utilization of more methods would result in more time and possibly expense, which would only be accomplished if teaching was considered to be very important. These analyses clearly ascertained that bachelor's departments placed a greater importance on quality teaching than other department levels, and the bachelor's departments typically employed more teaching evaluation methods to ensure accountability for quality teaching.

### **Chairperson Interview Data**

I made attempts to interview chairpersons from geography departments in all nine census divisions at each level of highest degree offered for a total of 27 interviews. I was able to complete interviews with 26 chairpersons from 25 of the 27 total department level-census region cohorts. I interviewed two chairpersons from departments in the same cohort because one U.S. state had not been represented in the faculty surveys. I

wanted information about teaching evaluation systems from every state, so I contacted this chairperson after I realized the state was not represented and, fortunately, the chair agreed to the interview. I was unable to schedule interviews with chairpersons from the doctoral-level South Atlantic departments and the bachelor's-level Middle Atlantic departments. The interviews I conducted represented 8 doctoral-level departments, 10 master's-level departments, and 8 bachelor's-level departments in 24 states. Most of the chairpersons were in charge of departments at public universities; only two interviewees chaired departments at private institutions.

I did not examine geographical differences because of the relative lack of significant differences among departments found in the quantitative analysis. I contemplated about departments with unionized faculty as a geographical trend due to the fact that the ten chairpersons indicating that their unionized faculty were located in union-friendly areas of the United States such as the states in the Northeast and the Great Lakes; however, I discovered as my interviews progressed that the power and influence of these unions regarding teaching evaluations varied considerably, even in the two foregoing regions.

The sample was also driven by the willingness of chairpersons to participate in the interview process. Initially, I attempted to arrange interviews by calling the chairpersons and scheduling accordingly; these attempts were somewhat successful. I was much more successful, however, scheduling interviews by contacting the chairpersons with an e-mail message briefly stating the purpose of my interview and my flexible interview schedule. I believe the greater success of the initial e-mail contact was

because the chairpersons being able to examine their schedules at their convenience as opposed to having to make a much quicker decision over the phone.

Although I used the interview framework (Appendix C) to structure the conversations, none of the interviews was exactly alike. Some chairpersons answered questions in the framework before I had the chance to pose them, but most of the difference was the result of follow-up questions I asked to provide clarification of or more information about a particular response. I adhered closely to the framework, though, because those questions were intended to gather the data needed for this study. The interviewees were all very forthcoming with responses to questions asked during the interviews; no chairperson declined to answer any of the questions from either the interview framework I developed or any follow-up questions I posed about their responses.

#### The Mandate for Evaluation of Teaching Effectiveness

I asked every chairperson first about the origin of the mandate for the formal teaching evaluation system. This question seemed a bit confusing at first to some of the chairpersons, so I explained the query by asking specifically whether the mandate came from the state government, the university, the college, or even the department. I was concerned this clarification may have given answers to the chairpersons they may not have otherwise given but I felt clarification was a more critical issue.

To the mandate question, department chairs gave answers with considerable variation (Table 21). Several chairpersons were unsure about the particular level at which the mandate actually originated, sometimes stating it could be a combination of



various entities. One chairperson honestly stated she had no idea about the origin of the mandate, although the answers she provided to subsequent questions seemed to indicate it was at least a university-level mandate.

**Table 21.** Chairperson Responses about the Mandate for Teaching Evaluation.

Mandate Level	Bachelor's Department Responses	Master's Department Responses	Doctoral Department Responses	Total Responses
Department	0	0	1	1
College	0	1	2	3
University/Provost	5	3	5	13
State University System	1	1	0	2
Board of Regents	0	1	1	2
Union	1	2	0	3
State Government	3	2	1	6
Accreditation Board	1	1	1	3

These responses represent a total greater than 26 due to some chairpersons responding that the mandate may have originated from multiple entities or institutions requiring the evaluation of teaching effectiveness simultaneously. For example, one chairperson replied that the mandate comes from the university as well as the state legislature and the accreditation board. Another chair explained that the university provides the mandate, but it may also come from the state university system and the faculty union.

A few trends emerged from the responses to this question. In almost every case, the mandate originated at some level higher than the department. Even the chairperson who indicated that the mandate for teaching evaluation was in the departmental operating document conceded that there was copious discussion about the need for teaching

evaluation at the university level and even the state legislature. Based on some of the uncertainty of the answers provided by chairpersons as to the absolute origin of the mandate, I found it difficult to make any clear statements or draw conclusions about the entities demanding accountability for teaching effectiveness other than those entities were almost always higher than the department level. When these responses were taken in conjunction with the next question in my interview framework, the mandate was often simply that evaluation of teaching effectiveness must be done with little or even no direction as to how it should be performed.

#### The Development of the Teaching Evaluation System and Instruments

Following the mandate question was an inquiry about the origins of the teaching evaluation system and the instruments of the system. Only two chairpersons responded the system and the instruments of that system did not have university, college, or department origins; one chairperson stated the system and instruments of the system, including what types of questions could be asked, were developed as part of the faculty union's collective bargaining agreement. Another chairperson reported there was a statewide adoption of the teaching evaluation system and instruments developed by the IDEA Center, a commercial producer of such systems. Overall, though, the development of the teaching evaluation systems seems much more localized, regardless of the origin of the mandate for teaching effectiveness to be evaluated.

The university provided teaching evaluation systems and instruments at 16 of the 26 departments whose chairpersons were interviewed. Within this group, considerable variability existed. Commercial forms, such as those developed by the IDEA Center, the

University of Washington, or the Educational Testing Service, were adopted without change by three universities. Another three universities had taken commercial systems such as these and adapted them for use on their campuses. The ten remaining universities that provided systems and instruments for teaching evaluation had developed them at their respective institutions.

However, even at universities that provided either commercial products or university-developed systems, departments often had considerable flexibility in terms of adding questions (either by department or by individual faculty adding questions to their course evaluations), selecting which specific instruments to use to evaluate their faculty, or even the option to develop their own system and instruments, as long as those departments adhered to university guidelines. One chairperson explained the university provides a student evaluation form that is widely used across the campus even without being required. Because of the familiarity of the form across campus, departments see its usage as helpful when promotion and tenure decisions reach university-level committees. Another chairperson detailed something similar in which departments are not required to use the University of Washington teaching evaluation products provided by the university, but the university encouraged them to do so for standardization purposes; faculty were still allowed to select which individual forms from the University of Washington teaching evaluation products were used in their particular classes.

Most chairpersons interviewed spoke of their student evaluation of teaching effectiveness systems in response to this question. The chairpersons who responded about peer evaluation of teaching effectiveness mentioned their departments had

developed these instruments, with one chairperson specifically emphasizing the peer evaluation forms had to be approved by the department's college.

None of the chairpersons stated explicitly the systems and their instruments were developed with any consultation of the extensive literature about the evaluation of teaching effectiveness. I was very careful not to mention literature consultation when asking about the development process because I did not want to influence the chairpersons' responses. The commercial systems and instruments have been developed in accordance with the recommendations of best practices from the literature. But none of the chairpersons at departments that developed their own systems and instruments or those at departments in which the university developed their own systems and instruments responded the literature had been consulted in the development process. Some chairpersons lacked the knowledge of the actual process that had been undertaken, either because the process took place before they were employed by their particular departments or simply not being informed about how the actual construction of the system and the instruments of the system had taken place.

### The Definition of Quality Teaching

Of the 26 chairpersons interviewed, only seven responded their departments had formal definitions of quality teaching. The remaining 19 chairs indicated there was no formal definition or that the definition of quality teaching was addressed inherently by the items on the teaching evaluation forms. Apparently, a very subjective nature undergirds the informal definitions of quality teaching at the departments that lacked formal ones. Chairpersons at those departments often used words and phrases like

“informal,” “subjective,” “philosophical,” “departmental culture,” “personal experience,” “you know it when you see it,” “if it sounds like good teaching, looks like good teaching, smells like good teaching, it’s good teaching,” suggesting remarkably qualitative approaches to what actually constitutes good teaching, even in cases where outcomes from teaching evaluation methods were quantitative data. Several chairpersons replied, because their departments were small, everyone has a sense of what is going on, who is doing a good job in the classroom, or who is engaging in quality teaching. As I listened to responses such as these, I understood departmental culture was often highly influential regarding the development of informal definitions of what constitutes quality teaching and what characteristics fit these definitions.

Most of the individual aspects of quality teaching provided by the chairpersons were those that occur frequently in definitions of quality teaching from the literature such as effectively communicating with students, providing clear expectations, providing relevant content, organizing courses in an appropriate manner, assessing students fairly to measure what they have learned, and stimulating student interest. Several chairpersons responded with aspects that are tied more closely to departmental or discipline-specific professional qualities, such as developing new courses or engaging students in fieldwork, individual projects, or research. A chairperson from a master’s-level department asserted specifically that involvement in the graduate program was one of the aspects examined when assessing the teaching quality of a faculty member.

Several chairpersons declared they examined whether faculty challenged students or maintained a certain degree of rigor in their courses without being overly punitive as a way of controlling for student evaluations of teaching effectiveness simply being

popularity contests. This approach was certainly one of the more interesting findings from the administration perspective of this issue of student evaluations as popularity contests. In fact, one chairperson related an anecdote about a faculty member who scored very high on student evaluations, yet the chair concluded students did not learn much in the faculty member's classes. Unique aspects such as these enhanced the variability among departments regarding what constitutes quality teaching.

#### The Value Placed on Quality Teaching by Department and Institution

I asked chairpersons to quantify the perceived value placed on quality teaching by their departments and institutions using the same equally stepped, five-anchor, Likert scale provided in the faculty surveys. Chairpersons rated the value placed on quality teaching in their departments no lower than a 3.5 out of 5.0; chairpersons rated the value placed on quality teaching in their institutions no lower than a 3. Chairs replied the perceived value placed on quality teaching by the institution was not higher than the perceived value placed on quality teaching by their departments. The mode for the responses to both questions was 5 (highly valued). Table 22 shows the mean scores from each value rating from the 26 chairpersons along with the averages for those same questions from the entire faculty survey sample.

Although the chair sample size was too small to make any rigorous quantitative analysis of these data useful, the patterns were identical to those observed in the faculty survey data. The chairs of bachelor's departments had the highest mean department value placed on quality teaching, the chairs of doctoral departments had the lowest, and the chairs of master's departments was in the middle. The general agreement between

faculty and chairs reinforces the view that bachelor's-level departments value teaching more than master's-level or doctoral departments.

**Table 22.** Means of Chairperson and Faculty Perceived Value Placed on Quality Teaching by Department and Institution.

	Bachelor's Department Chairpersons	Master's Department Chairpersons	Doctoral Department Chairpersons	Total Chairpersons	Total Faculty
Perceived Value Placed on Quality Teaching by Department	4.75	4.55	4.31	4.54	3.92
Perceived Value Placed on Quality Teaching by Institution	4.41	4.15	4.31	4.27	3.70

The mean scores of the chairperson responses were considerably higher than the overall faculty survey responses. Although I assured chairs interviewed their anonymity would be maintained, I still felt these results may have skewed higher due to the question being posed by interview as opposed to an online survey. Although a number of chairpersons were content to just answer with a number and move on to the next question, half of the chairs felt the need to explain the justification for their ratings for either one or both of the perceived value questions. I did not think any of the justifications that the chairpersons provided were defensive in nature, especially those who offered ratings lower than a 5.

## Remedies and Thresholds for Poor Teaching Evaluations

I asked chairpersons about actions concerning faculty members with poor teaching evaluations, specifically if there were remedies in place to address these circumstances, what these remedies may be, and what thresholds needed to be crossed to trigger those remedies. Having no experience with such issues during their time as chair, eight chairpersons responded with some uncertainty about how to answer. Interestingly, three chairpersons stated explicitly it was vital to address problems in the classroom in a timely manner because quality teaching was essential to the department's survival. The actual process for resolving such situations varied among the individual departments, although a number of departments shared common remedies.

Every chair responded chairperson or department head involvement in remedying poor teaching was imperative; the degree of that involvement, however, varied considerably. Eleven chairpersons answered struggling faculty members were handled on a subjective basis in terms of what remedies were suggested or actions taken. Four chairpersons said they would look at the nature of the courses in which the faculty member in question had received poor evaluations. These chairs emphasized challenging courses would raise fewer concerns, especially when taking the student grades in a course into account. Two of these chairpersons cautioned the decision to remedy may also be influenced by the faculty member having struggled in department courses that are part of the general education curriculum, and thus might attract students who have little interest in the material. The severity of the issue seemed to dictate the course of action the chair chose to take. One department, amazingly, had no remedies in place beyond simply encouraging the faculty member to improve, although continued poor teaching



performance could be an impediment to tenure/promotion decisions and merit raises. In another department, although remedies existed, the remedies were voluntary, and faculty members could not be required to take advantage of such remediation.

The options for remedying poor teaching performance available to each chairperson varied considerably, although, when examining the responses together, common options were available to a number of chairs (Table 23). The most common option, besides meetings with the chairperson to discuss improvement, was directing struggling faculty members to centers or workshops provided at the university level to help faculty persons improve their teaching; twenty chairpersons stated their universities had such centers or programs. In one department all new faculty were required to take a short course provided by that university's center for teaching.

Table 23 underscores departments, regardless of degree-granting level, employed similar remediation strategies for poor teaching, especially university programs/centers for teaching improvement and individual mentoring by skilled teachers. Four bachelor's departments reported poor teachers could also get assistance off campus at workshops and symposia. Three master's department responded "other," which explicitly included two systems that require faculty members to improve their teaching effectiveness or otherwise face dismissal. The state government mandated one of these systems; the university system authorized the other. The second procedure was a student evaluation committee comprised of undergraduate and graduate students who could make recommendations for improvement. The faculty union's collective bargaining agreement mandated this student evaluation committee. The student evaluation committee was not the sole remedy, however, simply part of the overall process.

**Table 23.** Frequency of Poor Teaching Remedies by Department Level.

	Bachelor's Department Chair Responses	Master's Department Chair Responses	Doctoral Department Chair Responses	Total Chair Responses
Department Committee	1	3	1	5
University Programs/Centers	7	6	7	20
Mentoring	3	3	4	10
Peer Evaluation	1	0	0	1
Chair Evaluation	1	0	0	1
Dean Involvement	2	2	0	4
Dismissal/Recommend Faculty Departure	1	1	2	4
Re-Assignment to Other Courses	1	1	0	2
Negative Merit Review	1	1	1	3
Assistance from Outside the University	4	0	1	5
Other	0	3	0	3

The mean number of remedies for poor teaching available to the bachelor's departments was higher than the master's or doctoral department means (Table 24).

**Table 24.** Mean Number of Remedies for Poor Teaching by Department Level.

	Total Number of Remedies	Mean Number of Remedies per Department
Bachelor's Departments (n = 8)	22	2.75
Master's Departments (n = 10)	20	2.00
Doctoral Departments (n=8)	16	2.00
Total	58	2.23

Every bachelor's department reported multiple remedies for poor teaching; five of the master's departments and three of the doctoral departments reported only one remedy. Some possible explanations for this trend exist. The chairs at bachelor's departments may have more options available. The reporting of remedies may have been higher due to the greater awareness chairs in bachelor's departments have of the available remedies for poor teaching. This higher awareness could possibly be because bachelors'-level chairs use the remedies more frequently. These reasons for a wider breadth of remediation could certainly be related to greater importance placed on teaching in bachelor's departments.

Some variability in the thresholds needed to be crossed to trigger these remedies was reported. I asked the chairpersons specifically about trends in the teaching evaluation data over time; below average performance on teaching evaluations provided data for specific thresholds. I defined a "long-term trend" as data from more than one academic semester or quarter and "immediate action" as data from one semester or quarter (Table 25).

Some chairpersons who required a long-term trend to have faculty persons enter into remediation explained the strategy resulted from faculty review timetables being in biennial increments. Some chairs asserted using teaching evaluation data from one course or even one semester/quarter would be unfair because not every semester/quarter is the same, because sometimes faculty can struggle in one course or one semester/quarter for a variety of reasons, such as teaching a course for the first time or just simply having a bad semester. Indeed, the most influential factor determining thresholds for remedying

poor teaching seemed to be the judgment of the chairs themselves. This propensity once again underscores the subjectivity of the remediation of poor teaching.

**Table 25.** Chair Responses by Department Level about Thresholds for Remedying Poor Teaching.

	Long-Term Trend Required	Immediate Action	Immediate Action for Tenure-Track; Longer Trend for Tenured	Increased Attention for Short-Term; Action for Long-Term Trend	None (Remedies are Voluntary)
Bachelor's Departments	4	1	1	1	0
Master's Departments	6	1	2	1	0
Doctoral Departments	4	3	0	0	1
Total	14	5	3	2	1

#### Differences in Teaching Evaluation Systems between Nontenured and Tenured Faculty

Table 26 displays the frequency and type of differences in teaching evaluation systems between nontenured and tenured faculty by department type.

**Table 26.** Frequency and Type of Differences in Teaching Evaluation Systems for Nontenured and Tenured Faculty by Department Type.

	Bachelor's Departments	Master's Departments	Doctoral Departments	Total
Same Systems for both Nontenured and Tenured Faculty	6	5	4	15
Difference in Frequency of Evaluation for Nontenured and Tenured Faculty	1	4	3	8
Differences in Evaluation Methods for Nontenured and Tenured Faculty	1	1	1	3

Fifteen of the twenty-six chairpersons interviewed responded that no differences existed whatsoever between nontenured and tenured faculty in terms of the evaluation of teaching effectiveness. Although the teaching effectiveness evaluation data for nontenured faculty at some of these departments may be scrutinized more closely, the methods and frequency of those methods remained the same despite faculty rank.

In every case where either frequency of method or number of methods differed, nontenured faculty members were subjected to the increase in frequency or the number of methods of evaluating teaching effectiveness. A larger number of master's and doctoral departments more frequently evaluated nontenured faculty members' teaching effectiveness than did bachelor's departments. Several of these department chairs offered a rationale for these differences in which the intention to address any teaching issues early, with the assumption that once any problems were fixed, the teaching would be satisfactory from that point onward. The consistency of the evaluation systems in bachelor's departments, though, seems to suggest a desire in those departments to monitor continually and consistently teaching performance, no matter the rank. This lack of assumption about teaching effectiveness tied to faculty rank reinforced the higher value placed on quality teaching by bachelor's departments.

Only three chairpersons responded particular teaching evaluation methods were weighted more heavily in the overall evaluation of teaching effectiveness. In one department peer evaluation for nontenured faculty was typically given more weight because of concerns that the student evaluations might be popularity contests rather than true measures of effective teaching. Student evaluations, according to one chair, were weighted a little more than peer evaluations, letters from the chair, or letters from the

dean because of concerns about colleagues being reluctant to criticize each other harshly. This chair also expressed negative peer or chair evaluations would be much greater issues due to the concerns. Chairperson evaluations in one department carried greater weight than student, peer, or external evaluations in the overall evaluation of nontenured faculty.

### Teaching as Part of the Total Faculty Workload

My next interview question addressed how teaching was included as part of the overall faculty workload. Three broad categories of responses unfolded for departments: (1) prescribed workload percentages for all faculty in teaching, scholarship, and service; (2) no prescribed workload percentages for all faculty in teaching, scholarship, and service; and (3) individual faculty work plans for teaching, scholarship, and service. I examined trends in these three broad categories, particularly regarding the level of department by highest degree offered, with the working assumption being that doctoral departments would place more emphasis on research and bachelor's departments would accentuate teaching.

Considerable variability exists in the data from the fourteen departments that have prescribed workload percentages for teaching, scholarship, and service. Table 27 displays the mean, maximum, and minimum percentages assigned to teaching, scholarship, and service from the fourteen departments combined.

Of the four departments with a maximum of 60% of total faculty workload devoted to teaching, one was a bachelor's department and three were master's departments. Every department with a teaching workload percentage equal to or greater than 50 percent was either a bachelor's or master's department; no doctoral department

had more than 40 percent of the total faculty workload assigned to teaching. The two departments representing the minimum percentage assigned to teaching included a master's department in which equal percentages were assigned to teaching, scholarship, and service (also resulting in this department having the maximum for service workload percentage), and a doctoral department. The two departments representing the maximum percentage assigned to scholarship were a doctoral department and, oddly, a bachelor's department. This bachelor's department chair did not provide a reason for this high percentage assigned to scholarship. The one department with only 20% of the faculty workload represented by scholarship was a bachelor's department. The percentage for service was assigned the lowest number or was tied for the lowest number in every department. The two departments having the minimum percentages were a bachelor's department and a master's department.

**Table 27.** Mean, Maximum, and Minimum Percentages for Teaching, Scholarship, and Service by Department Level.

		Teaching	Scholarship	Service
Bachelor's Departments	Mean	46	37	17
	Maximum	60	50	30
	Minimum	35	20	5
Master's Departments	Mean	52.8	33.1	14.1
	Maximum	60	40	33
	Minimum	33	25	5
Doctoral Departments	Mean	38.4	40.8	20.8
	Maximum	40	50	33
	Minimum	33	33	10
Total	Mean	45.7	36.9	17.3
	Maximum	60	50	33
	Minimum	33	20	5

Three departments employed the most frequently occurring percentage division of faculty workload, which was 40% teaching, 40% scholarship, and 20% service.

However, one chairperson at a doctoral department stressed this percentage is adjusted slightly for tenure-track faculty to 35% teaching, 45% scholarship, and 20% service. This adjustment was utilized to give opportunities for tenure-track faculty to establish their scholarship record more easily by reducing teaching responsibilities. As those tenure-track faculty gained tenure, their percentages become the 40-40-20 department standard.

Seven departments lacked any prescribed faculty workload percentages for teaching, scholarship, and service. One chairperson asserted, although there were no specific percentages, faculty members were expected to do sufficient work in all three categories. Another department had no specific percentages, but service was considered least important out of the three. One chairperson detailed the departmental system for faculty workload in these three categories with no percentages attached. For his department, faculty workload for these three categories changes as the faculty member's rank changes. At the assistant professor level prior to tenure, the faculty member must be excellent in teaching, meet university guidelines for scholarship, and provide departmental service. All three are important for gaining tenure, but teaching is the most important. After reaching the associate professor level, teaching is still expected to be excellent; any problems with teaching should have been fixed by this point, otherwise tenure would not have been granted. Scholarship is expected to continue meeting university standards, with the expectation that the research focus has moved beyond the faculty member's dissertation work by that time. Once a faculty member reaches the rank of professor, the service expectation changes; for example, professors are expected to provide university service by serving on university committees, becoming department



chair, and other more time-consuming service activities. This department's culture thus protects tenure-track faculty, particularly concerning service, which may take away from the development of quality teaching and scholarship.

The four remaining departments lacking specific faculty workload percentages showed trends that helped validate the assumption that scholarship is more valued at doctoral departments and teaching is more valued at bachelor's departments. The two chairpersons of doctoral departments underscored, although no specific percentages existed, scholarship was given more weight in terms of overall evaluation of faculty than either teaching or service. The two chairpersons of bachelor's departments confirmed the opposite for their departments, with teaching being more important than scholarship or service.

The remaining five departments have individual faculty work plans regarding teaching, scholarship, and service that are negotiated between the faculty member and chairperson on an annual basis. Interestingly, three of these departments were strong unionized systems in which the work plan agreement system had been negotiated and designed by the union. One chairperson simply responded that the yearly percentages for teaching, scholarship, and service can vary among faculty. Faculty work plans in two departments were typically 40% teaching, 40% scholarship, and 20% service with a few faculty having slight variability. One department had a faculty workload that was typically 65%-70% in teaching, 20%-25% in scholarship, and the remaining percentage in service.

The most interesting faculty work plan system was from a master's department with a strong union. This department could occasionally have faculty members devote an entire quarter in an academic year to scholarship or service with no teaching responsibilities; yet, all three categories must be represented in the yearly work plan agreement. Faculty rank in this department can have influence on these work plan agreements, with senior faculty possibly doing more service and tenure-track faculty possibly having more workload units in scholarship. On average, 32 out of the 45 workload units were related to instruction, but these units could include work with graduate students; not all of the instructional workload units were satisfied by classroom teaching.

#### Teaching Performance and Tenure/Promotion Decisions

The final interview question was about the relationship between teaching performance and tenure/promotion decisions. I followed up this broad inquiry with a hypothetical question about whether or not a faculty member who was a poor teacher but an outstanding researcher could receive promotion and tenure. Sixteen chairpersons underlined poor teachers would not receive promotion or tenure in their departments, even if those individuals were exemplary researchers, with many chairs exceptionally emphatic about this point. Two bachelor's department chairpersons and one at the master's-level stressed such individuals would not even be considered for promotion and tenure; those individuals would have been terminated during the tenure review process. Two other chairs at master's departments revealed poor teaching could possibly affect promotion and tenure but were reluctant to say that denial was certain. In three doctoral

departments, the converse was true as well: great teachers with poor scholarship records would not gain promotion and tenure.

Six chairpersons, surprisingly, indicated poor teachers with exemplary research records could gain promotion and tenure in their departments: one doctoral department, four master's departments, and two bachelor's departments. The doctoral department chairperson qualified his statement by saying the scholarship of the individual in question had to be of extremely high quality. One master's-level chair explained tenure was primarily based on teaching and promotion mainly grounded on research; this chair followed up by saying a poor teacher could get tenure but with the guarantee that significant improvement in teaching would follow. Another master's department chair clarified: if the faculty member in question was doing a large amount of graduate program work, specifically duties involving particular equipment or techniques in which the faculty member's skills were invaluable, then promotion and tenure could be achieved despite poor teaching; however, if the faculty member's teaching was primarily at the undergraduate level, there would be a problem with promotion and tenure.

In one bachelor's department either great scholarship or great teaching could gain promotion and tenure, even if the particular faculty member was weak in the other area. In another bachelor's department, a great teacher could gain promotion and tenure, even with a poor scholarship record, but the opposite could not happen. Even one doctoral department chairperson conceded a great teacher with just a mediocre scholarship record would probably be able to gain promotion and tenure in his department, although an average teacher with a great scholarship record would be able to do so as well.

Based on these results, most departments demand some sort of balance concerning faculty achievement in teaching and scholarship for promotion and tenure consideration. Although a few chairpersons conceded some flexibility in terms of the degree of faculty niche specialization in either teaching or scholarship, the prevailing attitude across departments is that teaching must be satisfactory to gain promotion and tenure. This finding also reinforces the responses from the interview questions about remedies for poor teaching and the greater focus on addressing any problems with teaching effectiveness in the tenure-track years. If poor teachers are allowed to remain ineffectual, then departments would simply allow these weak teachers to fail during promotion and tenure processes.

#### The First Qualitative Working Hypothesis

The responses from the chairperson interviews addressed the first working qualitative hypothesis: the teaching effectiveness evaluation methodology and utilization programs in geography departments in the United States have been developed without significant consultation of the literature pertaining to evaluation of teaching effectiveness, thus resulting in considerable variability throughout the discipline in regard to the application of the principles. Based on information from the chairperson interviews, I must accept this hypothesis for several reasons.

No chairperson stated the literature on teaching effectiveness evaluation was consulted in the development of the teaching evaluation system or instruments of the system. Although several departments used commercial products that were developed according from teaching evaluation literature, the large majority of universities,

departments, or other processes developed unique teaching assessment instruments that may or may not have included literature consultation. This finding may be the result of the lack of knowledge about how the development process took place, but the fact that no chairperson made any mention of the literature was striking.

The subjectivity in the analysis of the teaching effectiveness evaluation data was another reason for accepting the first qualitative working hypothesis. Although some of the subjectivity actually resulted in recommendations from the literature being taken into account, such as examining the types or levels of courses taught, I believe that much of the subjectivity was the result of most departments lacking a formal definition of quality teaching. No formal definition of quality teaching means no specific criteria or standards guide those who interpret the evaluation data. The variability of teaching effectiveness data sources, specifically within the same department where faculty of different ranks were evaluated using different or additional methods, also added to the subjective nature of the overall teaching evaluation systems.

This pervasive subjectivity is even more concerning given more than 60 percent of chairs acknowledged their departments would deny promotion and tenure based on poor teaching. Any legal issues that could arise from such denials would transcend the department level, particularly because higher university entities often mandate these systems and approve the departmental evaluation methods. The literature on teaching effectiveness evaluation systems, however, addresses these situations. The preponderance of evidence from the department chair interviews thus led to the acceptance of the first qualitative working hypothesis—most geography department

teaching effectiveness evaluation systems did not purposefully incorporate best practices recommended in the literature about these assessment systems.

### **Discussion of Faculty Survey Responses and Chairperson Interview Responses**

Although 26 departments were represented by the chairperson interviews, only 21 of those departments had corresponding responses from at least one faculty member to the faculty survey. I compared these faculty survey responses to the chairperson interview responses to discover whether any consistency exists among the faculty and administration understanding of teaching evaluation systems. Six questions from the faculty survey yielded corresponding responses from the chair interviews for the purposes of comparison:

- (1) Faculty values placed on quality teaching by the department.
- (2) Faculty values placed on quality teaching by the institution.
- (3) Methods used to evaluate teaching effectiveness for the responding faculty member by the department.
- (4) The weight given to teaching effectiveness in the department's tenure/promotion process.
- (5) Does the department have formal methods or programs for helping faculty improve low teaching evaluation scores?

(6) Does the institution have formal methods or programs for helping faculty improve low teaching evaluation scores?

By analyzing the faculty responses alongside the chair responses, I assigned a rating for each of the six questions based on the level of agreement between the faculty and the chair. I created a rubric comprised of four levels of agreement. These levels of agreement were defined as:

- Agree: many of the faculty agree with the chair completely.
- Slightly Agree: some of the faculty agree with the chair; the disagreements are not substantial.
- Slightly Disagree: many of the faculty disagree with the chair, or some disagree substantially.
- Disagree: many of the faculty disagree with the chair substantially.

After rating the departments' responses to each question, I constructed a matrix to compare the level of agreement with departments, among all departments, and among departments based on department level (Table 28). Among all 21 departments, the highest level of agreement between faculty and chairs was among the responses to the question about the existence of department remediation programs to help improve teaching effectiveness. This concurrence may be because most departments lacked such programs. Conversely, the level of agreement about the existence of university programs to help improve teaching effectiveness was quite low; many of the chairs responded that the universities had such programs, yet many faculty did not know if they existed or

actually stated they did not. These trends are probably the result of greater faculty familiarity with the department as opposed to the university.

**Table 28.** Matrix of Agreement between Faculty and Chairs\*.

Dept.	# Fac. Resp.	Value of Teach. -Dept.	Value of Teach. -Inst.	Teach. Eval. Methods	% of Teach. in T/P Process	Dept. Remedy Programs	Univ. Remedy Programs
Bachelor's Depts.							
1	2	SD	SD	A	A	SA	SA
2	1	A	SD	A	D	A	SD
3	5	D	D	D	D	SD	SA
4	3	A	SA	D	A	A	SA
5	1	SA	SA	A	SD	A	D
6	4	A	A	D	A	A	D
7	1	A	A	SA	SA	A	D
Master's Depts.							
1	1	SD	D	A	SD	A	SD
2	5	SA	SA	A	SA	A	A
3	2	A	A	SA	SA	A	A
4	1	SA	SA	D	D	A	A
5	8	SA	SA	SA	SA	A	D
6	2	SD	SA	SD	D	A	SD
7	2	A	A	A	SA	A	SA
8	2	D	D	SA	D	A	SD
Doctoral Depts.							
1	3	D	D	D	SD	A	D
2	4	D	D	SA	SA	SD	SD
3	3	A	A	A	SA	A	A
4	5	SD	SD	A	D	A	SD
5	3	A	A	SD	D	A	SD
6	2	D	D	SA	SA	SD	SD

\*A=Agree; SA=Slightly Agree; SD=Slightly Disagree; D=Disagree

The lowest level of agreement between faculty and chairs was among the responses to the question about the weighting of teaching in the tenure/promotion process. A number of chairs responded their departments had no formal percentages.



However, numerous faculty persons at the same departments replied with actual percentages. I tried to avoid such reporting errors by providing the “I don’t know” response, so this outcome may be a result of a number of faculty members being unaware about how much weight teaching actually carries in the tenure/promotion process.

In this analysis of agreement of faculty and chairs, differences also arose based on department level. Notwithstanding department 3 that expressed near-perfect agreement between faculty and chair, the doctoral departments (2.67 agreements per department), exhibited a lower degree of agreement between chairs and faculty than either the bachelor’s (3.71), except for the faculty in department 3 that almost completely disagreed with its chair, or master’s (4.13) departments. The trend of dissonance in doctoral departments implies that their faculties have a relative lower awareness of the nature of teaching evaluation systems and the use of the data produced by those systems. This lower faculty understanding could be a significant result of the lack of agreement in doctoral departments among the questions pertaining to the value placed on quality teaching by department and by institution, which was markedly more heightened in bachelor’s and master’s departments.

### The Second Qualitative Working Hypothesis

These comparative analyses were performed to address the second qualitative working hypothesis, which stated that agreement exists among faculty and department chairs/heads about the characteristics of the evaluation of teaching effectiveness methodology, utilization, and the value placed upon teaching by geography departments in United States and the value placed upon teaching by the institutions in which those

departments are located. From the foregoing analysis, I rejected this hypothesis based on the considerable amount of variability observed in many of the 21 departments.

The variation among the responses to the questions about the perceived value placed on quality teaching by department and institution was understandable in some ways, given the subjective nature of those questions. Although I made attempts to understand or even explain some of the variation observed among the other categories of responses, some of the inconsistencies defied explanation. I deferred to the chairperson responses as the ultimate authority on the composition of each department's teaching evaluation system and usage of teaching evaluation data. However, I was careful not to suggest any particular information in my structured interview questions or in my follow-up questions with the chairs because I did not want to unduly influence any responses. I doubt, though, that this would account for all of the variation observed.

Most of the considerable variation among the faculty members' responses could not be explained by differences in the teaching evaluation system according to faculty rank. I asked every chairperson to specifically address any differences in teaching evaluation methods or data utilization according to faculty rank to try to account for such differences. In some cases, such differences explained part of the variation. Error resulting from selecting responses accidentally or skipping responses inadvertently could also account for some of the variation. Given the total number of faculty responding to the survey, I doubt that error accounts for much of the variation. Overall, I deemed the amount of discrepancies to be large enough that no significant agreement between faculty members and chairpersons existed regarding teaching evaluation systems and the utilization of the data collected by those systems.

### **Discussion of Error and Bias**

The faculty members were surveyed using an online instrument that required faculty members to select their responses by clicking on them. On occasions faculty may have inadvertently selected answers they did not intend to or other types of user error. Indeed, the nature of the instrument itself may have caused some faculty members to decline to respond. Some faculty members expressed being confused about the questions pertaining to which teaching evaluation methods were used for formative and summative decisions. Although I defined both of these terms in the respective questions pertaining to each, this confusion may have affected the results, especially in terms of the number of “I don’t know” or “none” answers.

Errors may also exist in my qualitative analysis. The data were drawn from interviews with chairpersons; time constraints prevented me from interviewing the chairpersons at all 108 departments included in the total sample. Findings may have varied if different chairpersons, or if all chairpersons, had been interviewed. Also, not every chairperson interview had corresponding faculty survey responses, which may have affected the findings of the comparative analysis.

The chairpersons also had to agree to participate in the study. The chairpersons that participated may have had a particular interest in teaching or were exceptionally proud of the job being done regarding teaching in their departments, which may have skewed the results of the perceived value questions in particular. The interview format itself may have biased the results of those questions as well. Some of the chairpersons may have been reluctant to state low perceived values despite my guarantees about

anonymity. Moreover, I tried to control for any personal influence I might have had on the chairperson responses during the interviews. I was very careful when asking follow-up questions because of my knowledge of the recommendations for best practices in teaching evaluation in the literature.

Beyond these possible sampling errors, the statistical analyses were set within a 95% confidence interval that inherently produced error. Nevertheless, I worked diligently throughout this research to identify, understand, and manage possible error and bias.

## CHAPTER V

### CONCLUSIONS AND FUTURE RESEARCH

#### Conclusions

One of the important aspects of this study is the mixed-method examination of the teaching evaluation systems in geography departments in the United States. The quantitative analysis examined the geographical and department-level differences in the faculty survey responses from 101 departments regarding the perceived value placed on quality teaching by institution and by the department, and the total number of teaching evaluation methods. The qualitative analysis explored the interview responses gathered from 26 geography department chairs to discover any consistency among departments concerning aspects of the teaching evaluation systems and the usage of the data. The qualitative analysis also included a comparison between the faculty survey responses and chair interviews at 21 geography departments to determine the level of agreement between faculty members and chairs regarding the value placed on quality teaching by institution and department, aspects of the teaching evaluation systems, and the usage of the teaching assessment data.

The research was grounded within six quantitative working hypotheses and two qualitative working hypotheses.

### Quantitative Working Hypotheses

(1) No significant geographical differences according to the U.S. Census Divisions exist in the value placed on quality teaching by geography departments in the United States.

(2) No significant geographical differences according to the U.S. Census Divisions exist in the value placed on quality teaching by geography departments' institutions in the United States.

(3) No significant geographical differences according to the U.S. Census Divisions exist in regard to the total number of evaluation methods.

(4) No significant differences according to type of department based on the highest degree offered exist in the value placed on quality teaching by geography departments' institutions in the United States.

(5) No significant differences according to type of department based on the highest degree offered exist in the value placed on quality teaching by geography departments in the United States.

(6) No significant differences according to type of department based on the highest degree offered exist in the total number of teaching evaluation methods utilized by geography departments in the United States.

### Qualitative Working Hypotheses

(1) The teaching effectiveness evaluation methodology and utilization programs in geography departments in the United States have been developed without significant consultation of the literature pertaining to evaluation of teaching effectiveness, thus resulting in considerable variability throughout the discipline in regard to the application of the principles.

(2) Agreement exists among faculty and department chairs/heads about the characteristics of the evaluation of teaching effectiveness methodology, utilization, and the value placed upon teaching by geography departments in United States and the value placed upon teaching by the institutions in which those departments are located.

### Faculty Survey Data Analysis

This study confirmed there were no geographical differences among geography departments in the United States based on the perceived value placed on quality teaching by institution or department, resulting in the confirmation of the first two quantitative working hypotheses. The analysis of variance showed no significant differences among the nine U.S. Census divisions concerning these values. However, the analysis of variance did show significant differences among the nine U.S. Census divisions regarding the total number of teaching evaluation methods used by departments, resulting in the rejection of the third quantitative working hypothesis.

The chi-square tests resulted in the rejection of the final three quantitative working hypotheses. Geography departments were shown to differ significantly

according to department level in pertaining to the perceived value placed on teaching by institution and department, and the total number of teaching evaluation methods.

Bachelor's departments placed significantly higher value on quality teaching and expressed this higher value by employing substantially more teaching evaluation methods than master's or doctoral departments.

### Chair Interview Data Analysis

The qualitative analysis of the chair interview responses confirmed the trends revealed by the chi-square tests. Bachelor's departments placed a higher value on quality teaching, typically holding all faculty regardless of rank to those standards with their type and frequency of evaluation methods; sought to maintain that quality by providing more remedies to improve teaching effectiveness; designated substantial percentages of faculty workload for teaching activities; and rewarded quality teaching in promotion and tenure decisions more than master's or doctoral departments. However, this analysis displayed considerable variability among the individual departments, variability that often was not in accordance with the recommendations for best practices regarding teaching evaluation systems from the literature, resulting in the acceptance of the first qualitative working hypothesis. No two departments had the same system to address teaching effectiveness. This variability resulted from the considerable amount of autonomy many of these departments had in regard to the nature of their systems, despite the systems being mandated by higher-level entities such as the university, the state university system, the board of regents, or the state legislature. These systems lacked consistency even among departments with little or no say in how they evaluated teaching or used the evaluation data.



### Faculty Survey and Chair Interview Data Comparison

The comparison of faculty survey data and chair interview data from 21 geography departments showed a considerable lack of agreement within many departments regarding the value placed on quality teaching by institution and department, aspects of the teaching evaluation systems, and the usage of the assessment data, resulting in the rejection of the second qualitative working hypothesis. The responding faculty at every department failed to completely agree with the chair regarding all six common questions from the faculty survey and chair interviews about the perceived values placed on quality teaching by institution and department, the total number of teaching evaluation methods, the weighting of teaching in the tenure/promotion process, and the existence of department or university programs to improve teaching effectiveness. This finding pointed to an alarming paucity of faculty understanding about the nature of the teaching evaluation systems and the usage of the data from those systems. An examination of these departments by department level revealed higher levels of disagreement within doctoral departments than bachelor's or master's departments. I interpreted this lack of knowledge as a possible result of the lower value placed on quality teaching by doctoral departments. However, this generalization is tempered by the fact that within the department levels, the individual departments varied substantially.

### Hypotheses Outcomes and Theoretical Contributions

The goal of this study was to address the question posed in its problem statement: Are the methods and utilization of the evaluation of teaching effectiveness in geography

departments at colleges and universities in the United States consistent among the discipline along with the recommendation of good practice of teaching effectiveness evaluation found in the literature for evaluating teaching effectiveness for the discipline? Using the theoretical framework of organizational culture, my research hypotheses investigated the relationship between the hierarchy of subcultures (geographic location and department) and the teaching effectiveness evaluation systems in geography departments in the United States.

The quantitative analysis supported the hypotheses that geographical location did not have an effect on the values placed on quality teaching by institution or department. However, because of the responses from one of the nine U.S. Census divisions, this analysis did not support, albeit marginally, the hypothesis that geographical location does not have an effect on the total number of teaching evaluation methods. This result mildly supports McGuinness's (1999) idea that geographical location can influence higher education. The quantitative analysis, however, did not support the hypotheses that department type according to highest degree offered did not have an effect on the values placed on quality teaching by institution or department or the total number of teaching evaluation methods. Bachelor's departments and their institutions significantly valued teaching more than master's or doctoral departments, and expressed the higher value by developing and utilizing considerably more teaching evaluation methods. These findings support Schein's (2004) theory of organizational culture that the artifacts, in this case total number of teaching evaluation methods, are manifestations of the values placed on quality teaching in these department subcultures. Within the hierarchy of subcultures

developed at universities, as Hatch (2006) discussed theoretically, the department subculture dominated in regard to teaching evaluation systems.

The qualitative analysis supported the hypothesis that the teaching evaluation systems at geography departments in the United States had been developed without meaningful consultation of the literature pertaining to evaluation of teaching effectiveness, thus resulting in substantial variability throughout the discipline concerning the application of the principles. Bachelor's departments again significantly valued teaching more than master's or doctoral departments. This finding once again manifested itself in the development of such artifacts as more remediation methods for poor teaching and placing considerable weight on teaching in the tenure/promotion process that reflected those values. Hatch's (2006) explanation of the development of subcultures based on common interest or the result of frequent interaction explains the extensive variation among the all the individual departments and even departments within each level. These departments, especially those with considerable autonomy, have developed teaching evaluation systems that are expressions of the values individual departmental subcultures, regardless of the recommendations for best practices from the teaching evaluation literature. In some cases, these individual departments placed a high value on teaching because of the subculture of geography as a discipline, thus consistent with Middaugh's (2001) assertion that emphasis on teaching can be influenced by a discipline. Several chairs emphasized quality teaching was how their departments survived, so it was important to maintain high standards for teaching effectiveness. Nevertheless, the underlining conclusion is that most geography departments do not incorporate best practices from the literature as they devise teaching effectiveness evaluation systems,

which can, in turn, deleteriously affect the outcomes produced by the systems and the faculty assessed.

The qualitative analysis did not support the hypothesis that agreement exists among faculty and department chairs/heads about the characteristics of the evaluation of teaching effectiveness methodology, utilization, and the value placed upon teaching by geography departments in United States and the value placed upon teaching by the institutions in which those departments are located. The higher frequency of disagreement within doctoral departments was viewed as a result of lack of faculty understanding about the teaching evaluation systems, which may be caused from the lower values placed on quality teaching in the departments.

In sum, the methods and utilization of the evaluation of teaching effectiveness varied considerably throughout geography departments in the United States. One of the strongest influences on these department subcultures was the highest degree offered, which produced a schism between the bachelor's departments and graduate departments, particularly doctoral. This study provided a comprehensive examination of how geography departments evaluate teaching effectiveness and explained much of the existing variability.

### **Future Research**

Time constraints prevented an examination of the teaching evaluation system in every geography department in the United States. An investigation that utilizes more chair interviews with more questions, or a combination of survey and interview data

collection methods, may be useful in discovering more of the intricacies of the individual department teaching evaluation systems and the reasons why those unique qualities exist. This study also looked at teaching evaluation systems very broadly. Research that examines the instruments themselves or an in-depth analysis of the usage of the data collected by those instruments could prove enlightening.

This study showed that department level based on highest degree offered was a significant influence on the value placed on quality teaching by institution. A mixed-method study similar to this one using different departments at different levels at the same institution may be useful in terms of discovering the influence of the institutional culture on teaching evaluation systems. Moreover, a mixed-method study involving multiple departments at multiple public institutions in the same state may yield valuable findings regarding the influence of state governments that demand accountability for teaching performance.

Finally, incorporating a much larger database of information from geography departments across the country, a long-term goal should be to collaborate with colleagues who have expertise in educational assessment to try to construct a valid and reliable teaching evaluation model for geography, in particular, which is solidly based on the literature about teaching effectiveness assessment.

## APPENDIX A

### SAMPLE DEPARTMENTS FOR FACULTY SURVEY DISTRIBUTION

<b>Division</b>	<b>Doctorate</b>	<b>Master's</b>	<b>Bachelor's</b>
<b>1</b> <b>(CT, ME, MA, NH, RI, VT)</b>	Connecticut Clark Boston University Massachusetts	Central Connecticut St. Bridgewater St. Salem State Vermont	Southern Maine New Hampshire Rhode Island College Middlebury College
<b>2</b> <b>(NJ, NY, PA)</b>	Rutgers Syracuse SUNY-Buffalo Penn State	Montclair State SUNY-Albany Shippensburg Temple	Rowan Villanova Hofstra Colgate
<b>3</b> <b>(IN, IL, MI, OH, WI)</b>	Indiana S. Illinois-Carbondale Ohio State Wisconsin	Northern Illinois Toledo Eastern Michigan IUPUI	DePaul Youngstown St. Northern Michigan Wisc.-Eau Claire
<b>4</b> <b>(IA, KS, MN, MO, NE, ND, SD)</b>	Iowa Kansas Missouri-Kansas City Nebraska	North Dakota South Dakota State Missouri-Columbia Northern Iowa	Minn.-Duluth Gustavus Adolphus Macalester Central Missouri
<b>5</b> <b>(DE, DC, FL, GA, MD, NC, SC, VA, WV)</b>	Delaware Florida State North Carolina South Carolina	Marshall George Washington Appalachian State Florida Atlantic	Frostburg State Georgia Southern Old Dominion James Madison
<b>6</b> <b>(AL, KY, MS, TN)</b>	Kentucky Tennessee Southern Mississippi Georgia*	Auburn Alabama Eastern Kentucky Western Kentucky	Middle Tenn. State North Alabama Jacksonville State Louisville
<b>7</b> <b>(AR, LA, OK, TX)</b>	Louisiana State Oklahoma Oklahoma State Texas	New Orleans Arkansas North Texas New Mexico State*	Central Oklahoma East Central Texas Tech Sam Houston State
<b>8</b> <b>(AZ, CO, ID, NM, MT, UT, NV, WY)</b>	Arizona Colorado Idaho Nevada	Wyoming Montana New Mexico Northern Arizona	Brigham Young Northern Colorado Weber State Fresno State*
<b>9</b> <b>(AK, CA, HI, OR, WA)</b>	California-Berkeley Oregon Oregon State Hawaii	Central Washington Western Washington Cal. St.-Chico Cal. St.-Fullerton	Cal. St.-Sacramento Alaska-Fairbanks Eastern Washington Cal. St.-Stanislaus

\*Departments added from neighboring divisions to make up for shortfalls within divisions for those particular degree categories.

## APPENDIX B

### TEACHING EVALUATION SURVEY FOR GEOGRAPHY FACULTY

**1. What is your position? (select one)**

- Professor
- Associate Professor
- Assistant Professor
- Lecturer/Instructor
- Part-Time/Adjunct

**2. How many courses do you teach in a typical semester? (select one)**

- Zero    One    Two    Three    Four    Five or More

**3. Which option best describes your typical teaching load in terms of class size each semester? For this question, use the following standards: 30 or less = small; 31-100 = medium; more than 100 =large. (Select one)**

- All Large Classes
- Large and Medium Classes
- Large, Medium, and Small Classes
- Large and Small Classes
- All Medium Classes
- Medium and Small Classes
- All Small Classes

**4. In your opinion, to what degree is quality teaching valued by your university? (select one)**

Highly Unvalued      Unvalued      Neutral      Valued      Highly Valued

**5. In your opinion, to what degree is quality teaching valued by your department? (select one)**

Highly Unvalued      Unvalued      Neutral      Valued      Highly Valued

**6. Do the teaching effectiveness evaluation instruments gather quantitative data (e.g. Likert scale ratings), qualitative data (free-form written responses from evaluators), or a combination of the two types of data? (select one)**

Quantitative      Qualitative      Both

**7. Are faculty in your department rated (scored) in the teaching effectiveness evaluation process? (select one)**

Yes      No      Not certain/don't know

**8. Are faculty in your department ranked according to teaching effectiveness evaluation scores? (select one)**

Yes      No      Not sure/don't know

**9. Which of the following methods are used by your department to evaluate the teaching effectiveness of faculty (select all that apply)?**

Student evaluations	Self-evaluations
Peer evaluations	Videos
Chair evaluations	Student Interviews
Dean evaluations	Exit/Alumni evaluations
External expert evaluations	Learning outcome measures
Teaching portfolios	Teaching awards



**10. Which of the methods you indicated above were developed by your department to evaluate the teaching effectiveness of faculty (select all that apply)?**

Student evaluations	Videos
Peer evaluations	Student Interviews
Chair evaluations	Exit/alumni evaluations
Dean evaluations	Learning outcome measures
External expert evaluations	Teaching awards
Teaching portfolios	Don't Know
Formal self-evaluations	None

**11. To your knowledge, were any of these methods developed using the recommendations found in the literature pertaining to the evaluation of teaching effectiveness? (select one)**

Yes      No      Not certain/don't know

**12. Which of the methods used by your department to evaluate the teaching effectiveness of faculty are intended to provide information for formative decisions to be made regarding teaching practice? (*Formative decisions are those made from data gathered and shared with the teacher while a course is in progress, and are intended to improve teaching effectiveness while the course is in progress.*) (select all that apply)**

Student evaluations	Videos
Peer evaluations	Student Interviews
Chair evaluations	Exit/alumni evaluations
Dean evaluations	Learning outcome measures
External expert evaluations	Teaching awards
Teaching portfolios	Don't Know
Formal self-evaluations	None

**13. Which of the methods used by your department to evaluate the teaching effectiveness of faculty are intended to provide data for summative decisions to be made regarding teaching practice? (*Summative decisions are those made from data that is shared with the teacher after a course has been completed and are intended to help make decisions such as retention/dismissal, promotion, tenure, and merit.*) (select all that apply)**

Student evaluations	Videos
Peer evaluations	Student Interviews
Chair evaluations	Exit/alumni evaluations
Dean evaluations	Learning outcome measures
External expert evaluations	Teaching awards
Teaching portfolios	Don't Know
Formal self-evaluations	None

**14. Are there defined thresholds that faculty need to meet in order to satisfy department requirements for the level of teaching effectiveness? (select one)**

Yes      No      Not certain/don't know

**15. How much weight is given to teaching effectiveness in the promotion/tenure process in your department? (select one)**

0%      1-20%      21-40%      41-60%      61-89%      81-100%      Don't Know

**16. How much weight is given to teaching effectiveness in the merit process in your department? (select one)**

0%      1-20%      21-40%      41-60%      61-89%      81-100%      Don't Know

**17. Does your department have formal methods or programs for helping faculty improve low teaching effectiveness evaluation scores?**

Yes      No      Not certain/don't know

**18. Does your institution have formal methods or programs for helping faculty improve low teaching effectiveness evaluation scores?**

Yes      No      Not certain/don't know

## APPENDIX C

### INTERVIEW QUESTIONS FOR DEPARTMENT CHAIRS/HEADS

1. From where did the mandate for the formal evaluation system at your department originate?
2. From where did the resources/ideas used to develop the system and the instruments of that system originate?
3. How were these resources/ideas used to develop the system and/or instruments?
4. How is “quality teaching” defined by your department?
5. What aspects of teaching fit that definition?
6. Using a scale from 1 to 5, with 1 representing “Highly unvalued”, 2 representing “Unvalued”, 3 representing “Neutral”, 4 representing “Valued” and 5 representing “Highly valued” how would you rate the value placed on quality teaching by your department?
7. Using the same scale, how would you rate the value placed on quality teaching by your institution?
8. If a faculty member were to get consistently low evaluations, are there remedies in place to address these circumstances?
9. What is the threshold for requiring these remedies to improve teaching effectiveness?
10. What are those remedies for improving teaching effectiveness?
11. Do the evaluation methods differ between tenure-track and tenured faculty?
12. If so, how do these evaluation methods differ for each type?
13. What types of evaluation methods are used for each group?

14. How is the overall evaluation calculated; specifically, what is the weighting in a teaching evaluation score for each type of evaluation data for each group as a percentage of the entire evaluation?
15. If there are differences in weighting, why do they differ?
16. What is the weighting (in percentages) of teaching, scholarship, and service in terms of rank and importance as used by your department to evaluate the overall performance of each faculty member?
17. What is the relationship between teaching performance and promotion/merit/tenure decisions in your department?

## WORKS CITED

- Alexander, F. King. 2000. The changing face of accountability: Monitoring and assessing institutional performance in higher education. *The Journal of Higher Education* 411-31.
- Altbach, Philip G. 1999a. Harsh realities: The professoriate faces a new century. In *American higher education in the twenty-first century*, ed. Philip G. Altbach, Robert O. Berdahl, and Patricia J. Gumport, 271-297. Baltimore: Johns Hopkins University Press.
- Altbach, Philip G. 1999b. Patterns in higher education development. In *American higher education in the twenty-first century*, ed. Philip G. Altbach, Robert O. Berdahl, and Patricia J. Gumport, 15-37. Baltimore: Johns Hopkins University Press.
- Arreola, Raoul A. 1984. Evaluation of faculty performance: Key issues. In *Changing practices in faculty evaluation*, by Peter Seldin, 79-85. San Francisco: Jossey-Bass Publishers.
- Arreola, Raoul A. 2000. *Developing a comprehensive faculty evaluation system*. 2<sup>nd</sup> ed. Bolton, MA: Anker Publishing Company, Inc.
- Association of American Geographers. 2010. *AAG Guide to Geography Programs in the Americas 2009-2010*. Washington, DC: Association of American Geographers.
- Aubrecht, Judith D. 1984. Better faculty evaluation systems. In *Changing practices in faculty evaluation*, by Peter Seldin, 85-91. San Francisco: Jossey-Bass Publishers.
- Benenson, Gary. 1994. Closing the gap between teaching and learning: Formative evaluation in a freshman design course. *1994 Frontiers in Education Conference Proceedings, Twenty-Fourth Annual Conference* 739-43.

- Berk, Ronald A. 2006. *Thirteen strategies to measure college teaching*. Sterling, VA: Stylus Publishing, LLC.
- Berdahl, Robert O., Philip G. Altbach, and Patricia J. Gumpert. 1999. The contexts of American higher education. In *American higher education in the twenty-first century*, ed. Philip G. Altbach, Robert O. Berdahl, and Patricia J. Gumpert, 1-11. Baltimore: Johns Hopkins University Press.
- Berdahl, Robert O., and T.R. McConnell. 1999. Autonomy and accountability: Who controls academe? In *American higher education in the twenty-first century*, ed. Philip G. Altbach, Robert O. Berdahl, and Patricia J. Gumpert, 70-88. Baltimore: Johns Hopkins University Press.
- Birnbaum, Robert. 1988. *How colleges work: The cybernetics of academic organization and leadership*. San Francisco: Jossey-Bass Publishers.
- Bogue, E. Grady, and Kimberly Bingham Hall. 2003. *Quality and accountability in higher education: Improving policy, enhancing performance*. Westport, CT: Praeger Publishers.
- Cannon, Robert. 2001. Broadening the context for teaching evaluation. *New Directions in Teaching and Learning* 88(Winter): 87-97.
- Centra, John A. 1993. *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey-Bass Publishers.
- Chalkley, Brian, Eric J. Fournier, and A. David Hill. 2000. Geography teaching in higher education: Quality, assessment, and accountability. *Journal of Geography in Higher Education* 24(2): 238-45.
- Creswell, John W. 1998. *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage Publications.
- Danahy, Anne. 2011. Universities presidents warn budget cuts would have “ripple effect”. Centre-Daily Times.  
<http://www.centredaily.com/2011/03/16/2587501/universities-presidents-warn-budget.html> (accessed March 16, 2011).

- D'Andrea, Vaneeta-marie, and David Gosling. 2005. *Improving teaching and learning: A whole institution approach*. Berkshire, UK: Society for Research in Higher Education and Open University Press.
- Dressel, Paul. 1980. The nature and components of autonomy. *New Directions for Institutional Research* 7(2): 1-9.
- Estaville, Lawrence E., ed. 2010. *Geography Undergraduate Programs: Pathways to Success*. Online.  
<http://www.ncge.org/files/Geography%20Undergraduate%20Programs%20Revised%20Final%2012Jan11.pdf>. Accessed 1 April 2011.
- Fite, David. 2006. Using evaluation data for personnel decisions. In *Evaluating faculty performance: a practical guide to assessing teaching, research, and service*, by Peter Seldin and Associates, 181-200. Bolton, MA: Anker Publishing Company, Inc.
- Gaither, Gerald, Brian P. Nedwek, and John E. Neal. 1994. Measuring up: The promises and pitfalls of performance indicators in higher education. *ASHE-ERIC Higher Education Report No. 5*. Washington, DC: The George Washington University, Graduate School of Education and Human Development.
- Gates, Susan M., Catherine H. Augustine, Roger Benjamin, Tora K. Bikson, Tessa Kaganoff, Dina G. Levy, Joy S. Moini, and Ron W. Zimmer. 2002. Ensuring quality and productivity in higher education: An analysis of assessment practices. *ASHE-ERIC Higher Education Report Volume 29, No. 1*. San Francisco: Wiley Periodicals.
- Halonen, Jane S., and George B. Ellenberg. 2006. Teaching evaluation follies: Misperception and misbehavior in student evaluations of teachers. In *Evaluating faculty performance: a practical guide to assessing teaching, research, and service*, by Peter Seldin and Associates, 150-165. Bolton, MA: Anker Publishing Company, Inc.
- Hatch, Mary Jo. 2006. *Organization theory: Modern, symbolic, and postmodern perspectives*. 2<sup>nd</sup> ed. New York: Oxford University Press.



- Higgerson, Mary Lou. 2006. Building a climate for faculty evaluation that improves teaching. In *Evaluating faculty performance: a practical guide to assessing teaching, research, and service*, by Peter Seldin and Associates, 35-49. Bolton, MA: Anker Publishing Company, Inc.
- Jacobson, Robert L. 1992. Colleges face new pressure to increase faculty productivity. *Chronicle of Higher Education*, April 15.
- Kachigan, Sam. 1991. *Multivariate statistical analysis*. 2<sup>nd</sup> ed. New York: Radius Press.
- Knapper, Christopher. 2001. Broadening our approach to teaching evaluation. *New Directions in Teaching and Learning* 88(Winter):3-9.
- Magnusen, Karl O. 1987. Faculty evaluation, performance, and pay: Application and issues. *The Journal of Higher Education* 58(5): 516-529.
- McGuinness, Aims C. 1999. The states and higher education. In *American higher education in the twenty-first century*, ed. Philip G. Altbach, Robert O. Berdahl, and Patricia J. Gumpert, 183-215. Baltimore: Johns Hopkins University Press.
- Middaugh, Michael F. 2001. *Understanding faculty productivity*. San Francisco: Jossey-Bass Publishers.
- Millis, Barbara. 2006. Peer observations as a catalyst for faculty development. In *Evaluating faculty performance: a practical guide to assessing teaching, research, and service*, by Peter Seldin and Associates, 82-95. Bolton, MA: Anker Publishing Company, Inc.
- Moses, James, and Gary Boas. 1998. *Evaluating the college: Performance assessment in higher education*. New York: Primary Research Group, Inc.
- O'Neil, Robert M. 1999. Academic freedom: Past, present, and future. In *American higher education in the twenty-first century*, ed. Philip G. Altbach, Robert O. Berdahl, and Patricia J. Gumpert, 89-108. Baltimore: Johns Hopkins University Press.

- Pallett, William. 2006. Uses and abuses of student ratings. In *Evaluating faculty performance: a practical guide to assessing teaching, research, and service*, by Peter Seldin and Associates, 50-65. Bolton, MA: Anker Publishing Company, Inc.
- Rhodes, Frank H. T. 1999. The new university. In *Challenges facing higher education at the millennium*, ed. Werner Z. Hirsch and Luc E. Weber, 167-174. Phoenix: The American Council on Education and The Oryx Press.
- Schein, Edgar H. 2004. *Organizational culture and leadership*, 3<sup>rd</sup> ed. San Francisco: John Wiley and Sons, Inc.
- Seldin, Peter. 1984. *Changing practices in faculty evaluation*. San Francisco: Jossey-Bass Publishers.
- Seldin, Peter. 2006a. Building a successful evaluation program. In *Evaluating faculty performance: a practical guide to assessing teaching, research, and service*, by Peter Seldin and Associates, 1-19. Bolton, MA: Anker Publishing Company, Inc.
- Seldin, Peter. 2006b. Essential operating principles and key guidelines. In *Evaluating faculty performance: a practical guide to assessing teaching, research, and service*, by Peter Seldin and Associates, 1-19. Bolton, MA: Anker Publishing Company, Inc.
- Seldin, Peter. 2006c. Preface to *Evaluating faculty performance: a practical guide to assessing teaching, research, and service*, by Peter Seldin and Associates, xiii-xviii. Bolton, MA: Anker Publishing Company, Inc.
- Texas. Bicameral. 81st Legislature, regular sess. 2009. *H.B. 2504*. Accessed 16 March 2011; available from <http://www.legis.state.tx.us/tlodocs/81R/billtext/html/HB02504F.HTM>
- Tien, Chang-Lin. 1999. Future challenges facing American higher education. In *Challenges facing higher education at the millennium*, ed. Werner Z. Hirsch and Luc E. Weber, 161-166. Phoenix: The American Council on Education and The Oryx Press.

- U.S. Department of Education. 2006. *A test of leadership – Charting the future of U.S. higher education. A report of the commission appointed by Secretary of Education Margaret Spellings*. Online.  
<http://www2.ed.gov/about/bdscomm/list/hiedfuture/reports/final-report.pdf>.  
Accessed 25 March 2011.
- Weber, Luc E. 1999. Survey of the main challenges facing higher education at the millennium. In *Challenges facing higher education at the millennium*, ed. Werner Z. Hirsch and Luc E. Weber, 3-17. Phoenix: The American Council on Education and The Oryx Press.
- Zakrajsek, Todd. 2006. Using evaluation data to improve teaching effectiveness. In *Evaluating faculty performance: a practical guide to assessing teaching, research, and service*, by Peter Seldin and Associates, 166-180. Bolton, MA: Anker Publishing Company, Inc.
- Zusman, Ami. 1999. Issues facing higher education in the twenty-first century. In *American higher education in the twenty-first century*, ed. Philip G. Altbach, Robert O. Berdahl, and Patricia J. Gumport, 109-148. Baltimore: Johns Hopkins.

## VITA

Brian J. Cooper is the oldest of two children. He spent the first seventeen years of his life in Rock Hill, South Carolina. Brian began his undergraduate work at Clemson University in Clemson, South Carolina in 1996. After graduating *magna cum laude* from Clemson with a Bachelor of Science degree in Biological Sciences, he enrolled in the Master of Science program in Geography at Texas State University-San Marcos. After graduation with a Master of Science degree in Geography, Brian enrolled in the doctoral program at the Department of Geography at Texas State University-San Marcos to pursue a Ph.D. in Geographic Education.

Brian has taught several thousand students in three different geography courses during his eight year teaching career at Texas State University-San Marcos as a graduate student, adjunct lecturer, lecturer, and senior lecturer. He is the author of his own World Geography workbook, and has received numerous teaching honors both as a graduate student and as a full-time faculty member at Texas State University-San Marcos. Brian will continue his career in academia, focusing on teaching while continuing to develop his scholarship.

Permanent E-Mail Address: [bjcoope@gmail.com](mailto:bjcoope@gmail.com)

This dissertation was typed by Brian J. Cooper.