2008

Effective Teaching & Teaching Evaluation Practices: Canadian University Geography Departments

Susan Vajoczki
Wilfrid Laurier University

Effective Teaching & Teaching Evaluation Practices: Canadian University Geography Departments

Your file Votre référence
ISBN: 978-0-494-38008-6
Our file Notre référence
ISBN: 978-0-494-38008-6

Effective Teaching &
Teaching Evaluation Practices: Canadian University
Geography Departments

by

Susan Vajoczki

B.A., McMaster University, 1988

Hons. B.A., McMaster University, 1989

M.Sc., McMaster University, 1993

THESIS

Submitted to the Department of Geography and Environmental Studies/Faculty of Arts

in partial fulfilment of the requirements for

PhD

Wilfrid Laurier University

2008

DOCTOR OF PHILOSOPHY (2008)

(Geography)

Wilfrid Laurier University

Waterloo, Ontario

TITLE:

Effective Teaching & Teaching Evaluation Practices: Canadian University Geography Departments

AUTHOR: Susan Vajoczki

SUPERVISOR: Dr. Mary Louise Byrne

NUMBER OF PAGES: x, 195

ABSTRACT

In this thesis teaching evaluation practices in Canadian university geography departments are examined. The objective of this research is to identify good practices for teaching evaluation that can be applied within geography departments at Canadian universities and may be applicable to other departments and within other countries. In order to meet this goal a number of research questions were identified. These include:

1. What is effective teaching in higher education?;
2. What is effective teaching within the discipline of geography in higher education?;
3. How and why is teaching evaluated in higher education?;
4. What is the breadth of teaching evaluation practices currently used in geography departments within Canada?;
5. How are the results of teaching evaluations used to enhance teaching quality within the discipline of geography in higher education?;
6. How are the results of teaching evaluations used to reward teaching excellence within the discipline of geography in higher education?; and,
7. What are 'good' teaching evaluation practices within the discipline of geography in higher education?.

A thorough review of the literature resulted in the development of a conceptual framework for effective teaching and for effective teaching evaluation. Both of these frameworks were tested using empirical data. The empirical data were collected from a national-level survey of geography departments across Canada (n=10), oral interviews with chairs of Canadian geography departments (n=23) and oral interviews with individuals suggested by chairs of Canadian geography departments (n=11).

The research provided validation of the conceptual framework of good teaching. This framework identified eight parameters of good teaching: discipline knowledge, course organization, delivery of instruction, student/instructor interaction, assessment tasks, administration, professional development and skill development. The results from the research suggested that the original conceptual framework of effective teaching evaluation was too simplistic. It also demonstrated that effective teaching evaluation occurs within the demographics and culture of place which results in the creation of an environment of evaluation. In this environment of evaluation seven parameters of an effective teaching evaluation system are identified including: defining good teaching, operationalizing good teaching, defining purpose of teaching evaluation, using a multiplicity of tools, employing an iterative teaching evaluation cycle, including a system of awards and including a mechanism for teaching enhancement.

In the future these conceptual frameworks can be tested in a variety of other disciplines and within different countries. The primary stakeholder in this research was chairs of geography departments. There are a number of other stakeholders described in this research and the frameworks could be tested from their perspectives.

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank the many people who contributed to the successful completion of this thesis. First and foremost, I would like thank the survey participants for contributing their time and ideas to this thesis. I have had the opportunity to spend over twenty years in the university classroom as a student and as a teacher. I would like to extend my thanks to the thousands of students that I have had the opportunity to teach, to learn along with and to reflect with about good teaching practices.

A sincere thank-you is extended to my supervisor; Dr. Mary Louise Byrne, for her guidance, support, encouragement and willingness to supervise a thesis project in an area outside of coastal geomorphology. My committee members, Dr. Bob Sharpe, Dr. Philip Howarth and Dr. Erika Kustra provided both support and guidance. They asked all the right questions to make me think in more detail about the topic. I am grateful to Dr. Carolyn Eyles, Dr. Walter Peace, John Maclachlan and Krista Chomicki for reviewing survey drafts and providing support during the project. I am especially grateful to Carolyn who has provided years of mentorship, friendship and opportunities to reflect on teaching excellence. I would like to thank Dr. John Drake, my boss at the time, who planted the seed to work on a PhD. My current supervisor, Dr. Susan Elliott, who enabled me to complete this project both through the opportunities she provided me at work and by asking questions about progress and methods at critical stages throughout the project. I am forever grateful for her involvement in this project. I have also received regular motivation from Drs. Susan Watt and Hok Woo to see this project through to

completion. Their regular requests for progress reports has helped me maintain motivation during this project. For this motivation I am grateful. Over the past year I have had the opportunity to work with an outstanding research assistant, Leigh Ayton. Her attention to detail and ability to accurately and efficiently complete mundane tasks has made her an asset to this project. Thanks Leigh!

My parents Bill and Marjorie Cox have shown me that with dedication, sacrifice, courage, love and pride just about anything can be accomplished. From a very young age they instilled in me a belief about the power of education. Completion of this thesis would not have occurred without the support of my husband, Alex, who never complained about the hours spent on this project. My two children, Josh and Katie, provided me the foundation and stability I needed while I worked on this project. Although unknown to them they kept both my life and this thesis in perspective. Thanks also to Aaron Bertram for providing mental health breaks during this project. It was very convenient that Aaron was working in Australia while I finished the thesis so he was up in the middle of the night to discuss ideas.

I would like to dedicate this thesis to Dr. Brian McCann who piqued my interest in physical geography more than twenty years ago in a first year class that I had taken as an elective. Brian hired me to my first full-time job at McMaster University; he provided mentorship about effective teaching and continued to provide both guidance and friendship well into his retirement years. Although he is not here to see me finish this project I know that he is proud of its ultimate completion.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE:

Introduction

1.1 Research Context

There is a call, like never before, for champions of teaching in higher education. This is evident in the opportunities for promotion and career advancement based on excellence in teaching (Brown et al., 2002) and in declarations about teaching such as Smith's (1991 p. 63) statement, while Commissioner of the *Commission of Inquiry on Canadian University Education,* that:

> "teaching is seriously undervalued at Canadian universities and nothing less than a total re-commitment to it is required."

Boyer (1990) and Rice (1986) supported this emphasis on teaching and argued that the definition of scholarship should be broadened to include the 'scholarship of teaching.' The Rae Report (2005 p. 17) refers to:

> "a renewed commitment to something very basic: teaching excellence"

and recommends:

> "direct new investments towards teaching excellence and education innovation so that students have increased opportunities for meaningful contact with faculty, and better facilities and equipment" (Rae, 2005 p.30, 53).

One of the new investments described in the Report is the creation of a Council on Higher Education that will have as a key responsibility:

> "leading a renewed focus on the pre-eminence of teaching and teaching excellence at post-secondary institutions" (Rae, 2005 p. 51).

This Council has been created as the Higher Education Quality Council of Ontario through the Higher Education Quality Council of Ontario Act, 2005 (HEQCO Annual Report 2006/2007). The increased emphasis on teaching in higher education creates the need for comprehensive tools aimed at evaluating teaching effectiveness.

1.2 Research Objectives

The overall objective of this research is to identify good practices for teaching evaluation that can be applied within geography departments at Canadian universities. In order to meet this objective seven research questions were identified. These are:

1. What is effective teaching in higher education?
2. What is effective teaching within the discipline of geography in higher education?
3. How and why is teaching evaluated in higher education?
4. What is the breadth of teaching evaluation practices currently used in geography departments within Canada?
5. How are the results of teaching evaluations used to enhance teaching quality within the discipline of geography in higher education?
6. How are the results of teaching evaluations used to reward teaching excellence within the discipline of geography in higher education?

7. What are 'good' teaching evaluation practices within the discipline of geography in higher education?

At the outset of this work there is an expectation (i.e., hypothesis) that the research-intensiveness and the union status of the institution will influence how teaching is evaluated (i.e., question four above) and what is done with the results of teaching evaluations (i.e., question five and six above). If this proves to be the case, then it is anticipated that good teaching evaluation practices may differ between institutions depending on the research-intensiveness and union status of the individual institutions. Ideally, the outcome of this work will be the identification of good practices for teaching evaluation that can be applied to other disciplines and within other countries.

1.3 Organization of the Thesis

In Chapter One the context and relevance for conducting this research, as well as outlining the objectives of the research is described. This first chapter also describes the organizational structure of the thesis and provides descriptions for key terms. In Chapter Two a review of the literature on the characteristics of effective teaching within higher education and specifically effective teaching within university geography departments is provided The synthesis of this literature is presented through the development of a conceptual framework which identifies of the parameters of effective teaching within university level geography departments. In Chapter Three the reasons why teaching is evaluated and the existing methods for evaluating teaching along with a critique of these methods informed by the existing literature is discussed. This chapter concludes with the development of a conceptual framework and identification of the parameters that

contribute to effective teaching evaluation. In Chapter Four the research design employed to identify good teaching and good teaching evaluation practices in higher education geography departments is described. The chapter includes an approach and rationale for the design of the national level survey and for the selection of interview candidates. In this chapter the questions that were used to guide the open-ended interviews are provided. In Chapter Five an overview of good teaching as described by the research participants and the mapping of this description to the conceptual framework that was described in Chapter Two is provided. In Chapter Six an overview and assessment of teaching evaluation practices in Canadian university geography departments is provided. This overview and assessment is informed by the research participants. Chapter Six ends with a conclusion which maps good teaching evaluation practices described by the research participants to the conceptual framework described in Chapter Three. In Chapter Seven a summary of the conclusions of this work, the applicability of these results to other disciplines and to other countries and some suggestions for further work in this area is provided.

1.4 Terminology

A number of terms that will be used extensively throughout this document require defining. These terms are: teaching, learning, higher education, evaluation, student evaluations of teaching effectiveness, effective and good practices.

a. Teaching

The dictionary defines teaching, as,

> "any manner of imparting information or skills so that others may learn" (Merriam-Webster Dictionary Online, 2004).

There are three components in this definition of teaching. The first component, 'any manner' identifies that teaching is a multi-faceted activity that requires a broad range of competencies and does not occur in a single manner (Cashin, 1989). The second component, 'information or skills' identifies that teaching involves the imparting of two types of information (i.e., knowledge and skill). A third item, attitude is often thought to change as a result of learning and that it represents the outcome from the gain of knowledge and skill. The balance between these two items and attitude is an important aspect of teaching within geography (Abler, 1994). The third component, 'so that others may learn', identifies the ultimate goal of teaching, which is to facilitate learning in other people. Ultimately, teaching is any activity that manipulates a student's environment in order to facilitate learning or behavioural change. The breadth of what is effective teaching will be elaborated in Sections 2.1 and 2.2.

b. Learning

There are many definitions of learning. For this research a simple dictionary definition is being used. The dictionary defines learning as the process:

> "to gain knowledge [of] or understanding of or skill in by study" (Merriam-Webster Dictionary Online, 2004).

This definition identifies that there are three aspects of learning: knowledge gain, understanding gain, and skill acquisition. As mentioned earlier, knowledge gain and skill acquisition result in changes in attitude, which can often be measured as a change in understanding of a topic and are an integral part of teaching and learning within

geography. The definition also identifies that knowledge gain and skill development occur through study (i.e., practice) which requires engagement of the learner.

c. Higher Education:

Higher education in this document refers to post-secondary education within Canada that occurs at the university level. A number of the US studies, in particular, use the terms university, college and higher education interchangeably.

d. Evaluation

Evaluation, in the educational context, has been defined by Beeby (1977 p.68) as:

> "the systematic collection and interpretation of evidence, leading, as part of the process, to a judgment of value with a view to action."

Wolf (1987) identified four main elements to the definition of evaluation. The first element is that the data must be collected 'systematically'; that is, in a planned and thoughtful way. The second element is that there will be a process of 'interpretation of the evidence'. According to Wolf (1987) this is the part of the procedure in educational evaluation that is often overlooked. Wolf (1987) also argued that data are often collected and conclusions drawn without any analysis or interpretation of data occurring. For example, the conclusion that the educational system is failing is often drawn from the observation of high attrition rates (i.e., student dropouts), but there are many other legitimate reasons that may cause students to drop out. The third element in the definition of evaluation is the 'judgment of value' which involves the evaluator or group of evaluators using the collected information to make a judgment about the value of the evaluated item (e.g., teaching effectiveness, program, curriculum or institution). Alternatively, the judgment may involve combining this evaluation with others to make a

decision about future policy or action. The fourth element in the definition of evaluation is the 'view to action'. Evaluations in education are decision-oriented and intended to lead to improvement in student learning. If evaluations do not result in decisions (i.e., actions), the evaluation could likely be dispensed with and the evaluator's and instructor's time used more wisely (Wolf, 1987). The terms evaluation and assessment are used interchangeably in the teaching effectiveness literature.

e. Student Evaluation of Teaching Effectiveness

Student Evaluations of Teaching Effectiveness (SETE) are evaluations that are completed by students often towards the end of a course. Students are asked to evaluate the effectiveness of their instructor, and may be asked a series of other questions concerning their experience(s) in the course. The terms student rating and course evaluations are used interchangeable with SETE in the literature.

f. Effective

Effective for the purpose of this work means to produce a desired outcome. It is not necessarily associated with enhanced efficiency.

g. Good Practices

For the purposes of this paper, good practices are defined as those practices/methods that are most likely to facilitate effective student learning. These practices are purposefully not called best practices because of the recognition that good practices are constantly evolving and their identification is an on-going iterative process.

Teaching and learning can appear simple in definition but are considerably more complex. It is important to fully understand and appreciate that complexity prior to attempting to understand the teaching evaluation processes. In the next chapter some of

the complexities about what is good teaching will be explored.

CHAPTER TWO:

Effective Teaching

This Chapter will commence with a thorough discussion of the attributes of effective teaching in higher education. Once the attributes of effective teaching have been defined, there will be an exploration of the literature about effective teaching within the discipline of geography. The chapter will conclude with a discussion of a conceptual framework for effective teaching that will be tested in this research.

2.1 Attributes of Effective Teaching

An extensive literature of more than 15,000 papers, presentations and articles about teaching effectiveness was published from 1989 to 1998 and over 5,000 of these described effective teaching in higher education with 1,500+ involving empirical studies of teaching in higher education (Moore, 1999). A Google Scholar search in November 2007 identified 1,710,000 items on effective teaching and 1,430,000 on effective teaching in higher education. Since the mid 1990s, a meta analysis has not been completed on effective teaching in higher education. Although there has been considerable research conducted on teacher effectiveness, the primary question, what is effective teaching, still remains (Young and Shaw, 1999). Much of the research has focused on students'

perceptions of the attributes of effective teaching garnered from Student Evaluations of Teaching Effectiveness (SETE). The preconceived ideas of what defines 'good' and 'bad' teaching with which students arrive at a classroom is based on 'apprenticeship of observation' and on a continuum from effective to ineffective characteristics (Walls et al., 2002).

Since the 1950s the focus of research on effective teaching has been on examining the relationship between inputs to the teaching process and the resulting product. The teaching process includes: classroom atmosphere (both within the confines of the four walls of the traditional classroom and the experiences had in non-traditional classrooms; e.g., field experiences), teachers' behaviours, student learning activities, course organization and evaluation procedures. The product examines both end-of-course and long-term learning with consideration of the attitude change, knowledge and skill acquisition of the students (Braskamp et al., 1984). This body of research has had two foci. One focus has been on the discovery of the characteristics that are associated with effective teaching including the identification of attributes, traits and personalities of the effective teacher (Centra, 1994; Cohen 1981). The second focus has been on the identification of the relationship between the type of content (e.g., factual vs. skills) and the best method of delivery (e.g., traditional lecture vs. discussion). The outcome of research about these foci is the recognition that effective teaching:

- is a complex, multi-faceted construct (Marsh and Roche, 1997; Cashin, 1989; Marsh, 1987);
- results in positive changes in student learning (Ramsden, 1992);
- comprises multiple perspectives (Young and Shaw, 1999);
- consists of a suite of effective teaching skills and techniques that are discipline-specific (Crebbin, 1997; Ramsden, 1991;

Murray and Renaud, 1995; Shulman, 1993; Sullivan and Skanes, 1974); and,

- is accomplished by instructors who have different abilities, skills and preferences. They should not only identify their own strengths and weaknesses, they should be encouraged to use them (Braskamp et al., 1984).

Table 2.1 provides a summary outlining the breadth of teaching effectiveness attributes identified by different researchers. Following the table is an explanation of each researcher's methodology. The work is presented in chronological sequence which highlights the focus on determining the characteristics of good teaching during the 1980s and 1990s. Much of this work is synthesized in the seven principles for good practice in undergraduate education developed by Chickering and Gamson (1987) which is described later in this chapter. The literature that informs Table 2.1 is very broad and often contradictory. To ensure the credibility of the information contained within the Table a rigorous process to review the literature was employed. The researcher spent over 15 years reviewing and evaluating this literature, using a very broad approach to access both literature that was published in discipline-specific journals and in education journals. This literature focuses heavily on the results from student evaluations of teaching effectiveness. Meta-analyses which were primarily completed in the 1970s and 1980s were given particular emphasis for their ability to synthesis common meaning from the broad literature. Both cue cards and sticky notes were used to dissect common threads from the literature. Additionally, award-winning teachers were consulted (antidotal evidence) and asked to comment on the common parameters that were emerging.

Table 2.1: Summary of the attributes of effective teaching. The effective teaching attributes are numbered in the order in which the researcher has presented the information in the original text

Authors	Attribute
Murray, 1980	1. Classroom presentation 2. Course content 3. Course management 4. Extra-curricular teaching
Marsh, 1984	1. Learning/Value 2. Enthusiasm 3. Organization/Clarity 4. Group Interaction 5. Individual Rapport 6. Breadth of Coverage 7. Examinations/Grading 8. Assignments 9. Workload/Difficulty 10. Overall Course 11. Overall Instructor 12. Prior Subject Interest 13. Reason for taking (% indicating general interest) 14. Class Average Expected Grade 15. Workload/difficulty 16. Course Enrollment 17. % attendance on day evaluations administered

Feldman, 1976	1. Teacher's stimulation of interest in the course and the subject matter 2. Teacher's enthusiasm for subject or for teaching 3. Teacher's knowledge of the subject 4. Teacher's intellectual expansiveness and breadth of coverage 5. Teacher's preparation and organization of the course 6. Clarity and understandableness of presentation and explanations 7. Teacher's elocutionary skills 8. Teacher's sensitivity to, and concern with, class level and progress 9. Clarity of course objectives and requirements 10. Nature and value of the course material including its usefulness and relevance 11. Nature and usefulness of supplementary materials and teaching aids 12. Difficulty and workload of the course 13. Teacher's fairness and impartiality of evaluation of students; quality of exams 14. Classroom management 15. Nature, quality and frequency of feedback from teacher to students 16. Teacher's encouragement of questions and discussion, and openness to the opinions of others 17. Intellectual challenge and encouragement of independent thought 18. Teacher's concern and respect for students; friendliness of the teacher 19. Teacher's availability and helpfulness
Feldman, 1988	1. Knowledge of subject or discipline 2. Course preparation and organization 3. Clarity and understanding 4. Enthusiasm for subject teaching 5. Sensitivity to/concern with students' level and learning progress 6. Availability and helpfulness 7. Quality of examinations 8. Impartiality in evaluating students 9. Overall fairness to students
Cashin, 1989	1. Subject matter mastery 2. Curriculum development 3. Course design 4. Delivery of instruction 5. Assessment of instruction 6. Availability to students 7. Administrative requirements

Abler, 1994	1. Advising undergraduate students 2. Supervising undergraduate thesis 3. Curriculum and course development 4. Developing and implementing innovative teaching approaches 5. Establishing and supervising student internships 6. Adaptation and application of technology to assist curricula instruction
Knapper & Rogers 1994	1. Course and curriculum planning 2. Designing and marking assessment tasks 3. Supervising undergraduate and graduate thesis and practica 4. Interacting with students 5. Keeping abreast of developments in the field 6. Advising colleagues on teaching issues
Young and Shaw, 1999	1. Effective communication 2. Comfortable learning atmosphere 3. Concern for student learning 4. Student motivation 5. Course organization 6. Value of the course
Chalkley et al., 2000	1. Clear goals and objectives 2. Strong understanding and knowledge base in the discipline 3. Appropriate level of difficulty 4. Qualities that motivate & engage students 5. Clear organization 6. Use of learning resources 7. Increasing complex challenges 8. Student focused 9. Links to appropriate formative and summative assessment 10. Achieve deep rather than surficial learning

Murray (1980) identified four main categories of teaching in an attempt to establish a teaching evaluation scheme at the University of Queensland that measured the breadth of teaching (Table 2.1). Previous teaching evaluation schemes at the University of Queensland had not successfully measured breadth. Although the categories are not specific attributes they are included in the list to demonstrate the scope of what needs to be considered teaching.

Marsh (1984) analysed SETE data from 1364 courses and identified 17 attributes of effective teaching (Table 2.1). Marsh's (1984) intent in collecting these data was not to identify the attributes of effective teaching; rather, it was to assess the generalizability of SETE results (Section 3.3). Although the attributes of effective teaching are a secondary outcome of this research, it is a valuable data set as the data were collected from a wide variety of courses at different institutions and in different disciplines; thus, it encompasses a wide breadth of effective teaching attributes.

In a comprehensive analysis of teaching effectiveness research, Feldman (1976) reviewed 72 studies that identified effective teaching attributes from SETE. Of these studies, 49 considered the characteristics of 'ideal and best' college teachers and characteristics important to superior college teaching as perceived by college students and 23 considered the association between college students' overall evaluation of their teachers and their evaluation of specific characteristics of these teachers. In this research the attitudes, behaviours, and pedagogical practices that are most associated with effective college teaching were determined four ways:

- students creating their own list of characteristics;
- students responding to a pre-set list of characteristics;

- comparing specific items on SETEs to the global evaluation of the instructor; and,
- comparing specific items that most frequently combined with global evaluation items to form the highest loadings on the same factor in factor-analytic studies.

One of the major contributions of Feldman's (1976) work was the identification of 19 instructional dimensions that are associated with effective teaching (Table 2.1). The two characteristics that were consistently associated with superior college teaching using all four methods were stimulation of ***student interest*** *and* ***clarity***. A third characteristic, ***knowledge of subject matter***, was fairly consistently associated with superior teaching. Two other characteristics that emerged consistently as important to superior teaching were ***instructor preparation and organization*** of the class and ***instructor enthusiasm*** for the subject material. Interestingly, characteristics including friendliness (concern and respect for students), helpfulness (availability), openness to others' opinions (encouragement of class questions and discussion) and the instructor's ability as a facilitator are among the items most frequently identified when students were asked their perception of characteristics most associated with 'ideal or best' teachers but were of much less importance when students responded to a pre-set list of characteristics. This disagreement between student definition of teaching effectiveness dependent on, the presence or lack thereof, of a pre-set list of characteristics has been explained in the past as occurring because students perceive these characteristics as less important when they consider more specific, structured, salient characteristics in the pre-set list of characteristics (Newcombe et al., 1965, Brown 1965). In subsequent studies (e.g., Feldman, 1984), difficulty and workload of the course and classroom management were

dropped from the list and perceived outcome or impact of the instruction and instructor personal characteristics (i.e., personality) were added.

Further meta-analytic work by Feldman (1988) analyzing studies of attributes of teaching effectiveness from SETEs resulted in the list being refined from 19 characteristics to nine characteristics that both instructors and students identified as important attributes of effective teaching (Table 2.1).

Cashin (1989), an instructional developer, identified seven attributes of effective teaching (Table 2.1) based on an extension of the work by Arreola (1986; 1989) and Centra (1977).

Abler (1994) identified six attributes of effective teaching that are common to all disciplines (Table 2.1). These attributes are based on the researcher's experience as a geographer. Item 4 (i.e., developing and implementing innovative teaching approaches) in Abler's (1994) list of six attributes of effective teaching is problematic. An effective teacher might very well be more successful developing and implementing teaching approaches that have demonstrated success as opposed to developing and implementing innovative teaching approaches. Abler (1994) may have included this item because the instructor that incorporates innovative teaching approaches is likely more reflective on their own teaching and more abreast of developments in the fields of teaching and learning. These characteristics (i.e., reflection and currency) would benefit teaching effectiveness. A second reason why Abler included this item may be the definition that was being used for innovative. For the purpose of the current research 'innovative' implies something new and different as opposed to some research that defines innovative as new to the instructor.

Knapper and Rogers (1994) identified six attributes of effective teaching in their paper prepared at the request of the Ontario Council of University Affairs for the Task Force on Resource Allocation (Table 2.1). The attributes selected are based on the researchers' experience as instructors and instructional developers.

Young and Shaw (1999) demonstrated that effective teaching can be predicted with many fewer dimensions than identified by Feldman (1976). In their study, Young and Shaw (1999) surveyed 912 students (530 graduate students and 382 undergraduate students) at a medium-sized U.S. university (enrollment 11,000 students) and identified that as a group, effective communication, a comfortable learning atmosphere, concern for student learning, student motivation, and course organization were found to be highly related to effective teaching (Table 2.1). The unexpected predictor of teacher effectiveness in this research was the value of the course (Young and Shaw, 1999). This predictor emerged when graduate students were considered separately, undergraduate students separately, and when the groups were combined (Young and Shaw, 1999). The interpretation of this result may be that the effective teacher is one that demonstrates to students the value of the course content (Young and Shaw, 1999). Value for the purpose of Young and Shaw's (1999) was defined as relevancy. Students placed a high value in the course material if they perceived it is relevant; but the researchers did not define relevant to what. Teachers who could demonstrate the relevance (i.e., value) were consistently perceived as effective.

Chalkley et al. (2000) proposed that a more 'common sense' definition of effective teaching is required. Without providing a specific 'common sense' definition the researchers identified ten attributes of effective teaching (Table 2.1). It appears that

this list is drawn from the experience of the researchers in the classroom. This study is important to this paper because the researchers were geographers focusing on effective geography teaching in higher education.

The work summarized in Table 2.1 concentrates almost exclusively on teacher behaviour and student-teacher interaction (with the following exceptions: i)Young and Shaw (1999) point six, value of the course which was discussed earlier and ii) Chalkley et al. (2000) point ten, achieve deep learning rather than surficial learning). Deep learning is learning that focuses on the overall meaning of the material and encourages students to relate ideas together to construct new concepts, whereas, surface learning focuses on students acquiring content knowledge and tends to lead to students memorizing details (Ramsden, 1992). Much of the focus in work on teaching effectiveness since the 1990s suggests that good teaching encourages high quality student learning by active student engagement with the subject content and discourages surficial approaches to learning that are represented by the imitation approach which focuses on memorization and shallow learning (Ramsden, 1992). Educators are beginning to recognize that excessive testing and workloads create surface learners as opposed to deep learners. Ramsden and Entwistle (1981) studied several thousand students at UK universities and identified clear relationships between deep learning approaches and the type of teaching students experienced, identifying that an effective teacher must focus on the learning process to ensure that deep learning occurs.

Andrews et al. (1996) argued that not only must the focus be on the learning process, but also that effective teachers must not be under the assumption that students reach university with a thorough understanding of how to learn; rather it should be

understood that they enter post-secondary education with only a basic level of cognitive development for understanding the nature of knowledge and how to acquire it. Smith (2000) identified that the effective teacher is the one who recognizes that preferred learning styles will differ between students, that deep learning should be promoted, and that there is an emotional element to learning. Thus, to facilitate the greatest learning, the emotional element must be captured.

The work of numerous researchers in attempting to define the attributes of effective teaching has led The American Association of Higher Education, the Education Commission of the United States and the Johnson Foundation to review more than 50 years of research on education practice and to identify seven principles for good practice in undergraduate education:

- encourages student-faculty contact;
- encourages cooperation among students;
- encourages active learning;
- gives prompt feedback;
- emphasizes time on task;
- communicates high expectations; and,
- respects diverse talents and ways of learning (Chickering and Gamson, 1987).

These seven 'practices' have received widespread attention and support in higher education because they are a clearly identifiable and an easily understandable list of concise activities that instructors can aim to accomplish.

Schank (2002) provided clarity on the future direction of teaching and learning with the argument that knowledge is becoming increasingly accessible and that true ability in the future will not be measured by an individual's ability to possess a great knowledge but in the ability to know what questions to ask and how to ask questions.

Thus, the effective teacher is the one who can 'change' the students' abilities so that they are better able to formulate questions and subsequently answer these questions.

2.2 Effective Teaching in Geography

There is widespread agreement among researchers that, although some of the multi-dimensional aspects to teaching are common across the disciplines, a number of the criteria will be discipline-specific (Burkill, 2002; Elton, 1998; Murray and Renaud, 1995; Ramsden, 1991). Geography is no exception and has a number of criteria that must be considered when defining effective teaching within the discipline. These criteria include:

- preparing for and conducting numerous and extended field trips (Tricart, 1969; Gold et al., 1991; Abler, 1994; Cooke, 1998; Chalkley et al, 2000);
- the interdisciplinary nature of geography (i.e., blending science, social science and humanities as well as 'borrowing' methodologies from a wide variety of other disciplines and linking theoretical and applied aspects), (Tricart, 1969; Abler, 1994; Marantz and Warren, 1998; Farrington, 2000; Geography Benchmarking Group, 2000);
- teaching topics that require computer-assisted teaching and learning e.g., GIS (Gold et al, 1991; Chalkley et al., 2000); and,
- a large component of teaching that involves instruction of audiences beyond the traditional tuition-paying students due to the strong sense of community responsibility and outreach (i.e., a large civic responsibility) (Abler, 1994).

Gold et al. (1991) in their book, *Teaching Geography in Higher Education,* have identified three additional generic principles to augment the list prepared by Chickering and Gamson (1987):

- good practice evaluates itself – stressing the importance of reflective self evaluation in addition to other evaluations;
- good practice is clear about its aims and objectives; and,
- good practice consults the educational literature.

Although Gold et al. (1991) are writing about geography teaching, none of the items they add are specific to the discipline of geography.

Effective teaching within geography must meet both the general characteristics of effective teaching outlined (see Section 2.1) and the characteristics derived from the nature of the discipline (see Section 2.2).

2.3 Conceptual Framework – Effective Teaching

The thorough review of the literature presented in this chapter has permitted the researcher to develop a conceptual framework of effective teaching (Figure 2.1). This conceptual framework illustrates that effective teaching, is impacted by eight parameters: discipline knowledge, course organization, delivery of instruction, student/instructor interaction, assessment tasks, administration, professional development and skill development. These parameters do not occur in isolation and the arrows illustrate that they are all interconnected. One parameter in particular, skill development, includes a number of items that are discipline-specific, which in the case of geography, may include mapping, geographic information systems (GIS), and field techniques. In order to test it, the conceptual framework that has been developed through the review of the literature will be examined and tested through interviews with geography chairs, university administrators, staff in teaching and learning centres, and undergraduate students.

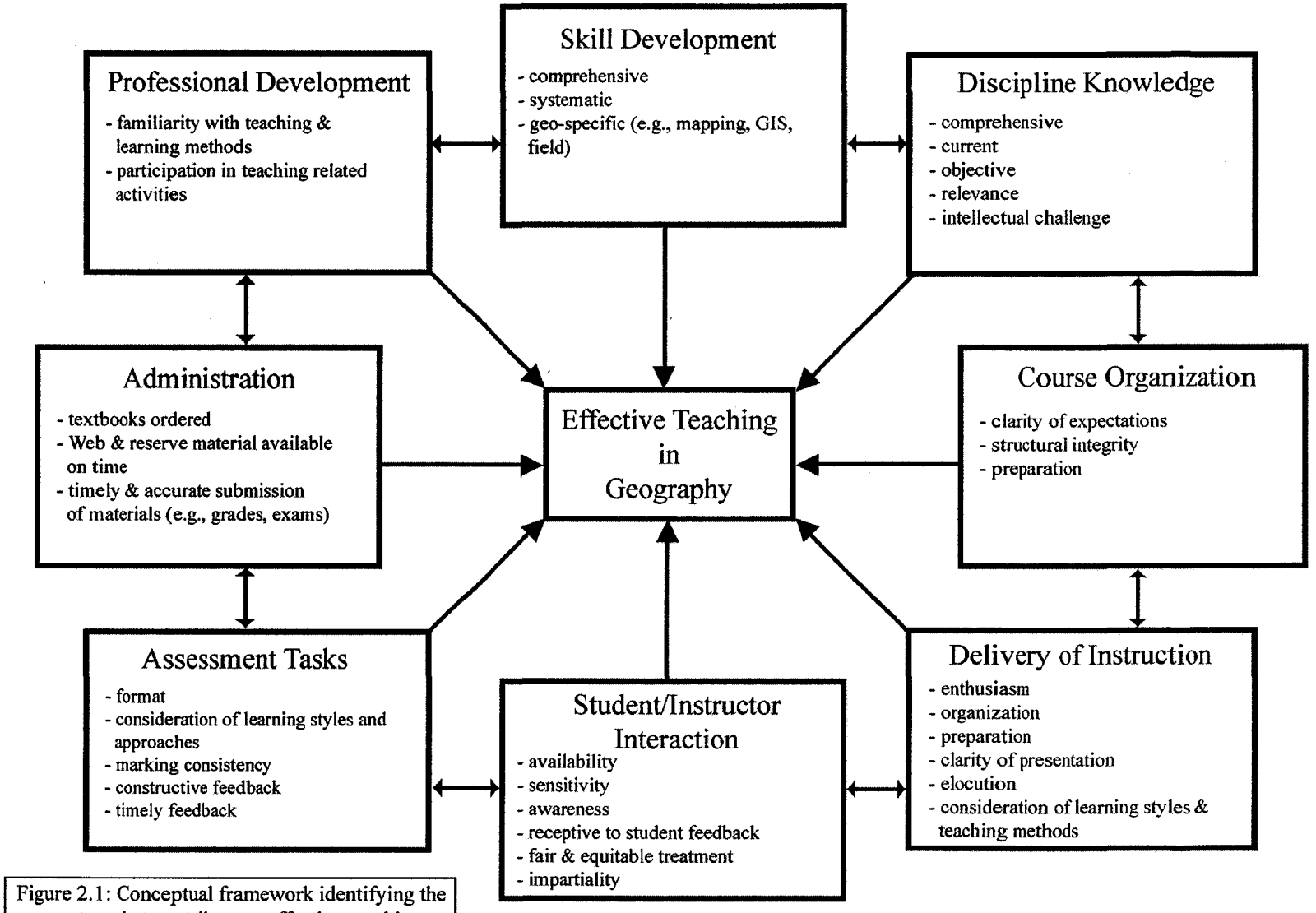


Figure 2.1: Conceptual framework identifying the parameters that contribute to effective teaching.

CHAPTER THREE:

Teaching Evaluation

This chapter will commence with a discussion about the current interest level in the teaching evaluation practice. This will be followed by identification of the stakeholders in the teaching evaluation process and a brief needs assessment. The goals and purposes of evaluating will then be described followed by a review of the literature about the elements of an effective teaching evaluation system and the possible tools to evaluate teaching. A discussion of the opportunities and constraints for effective teaching evaluation within geography will follow. The final section in this chapter will be a discussion of a conceptual framework for effective teaching evaluation that will be tested in the current research.

3.1 Interest in Teaching Effectiveness Assessments

There are several reasons for the recent interest in enhanced teaching quality. These include: higher participation rates, increased costs of education, public demand for accountability, and a need to develop practical skills.

Chalkley (1998) argued that one reason for the increased emphasis on evaluation in higher education is the increased participation rate of students in higher education. His argument is that with the greater participation rate there is a necessity for greater

emphasis on quality. This argument is supported with evidence from Statistics Canada showing a 25% increase in the number of individuals in the age bracket 15 years+ during the time period 1990-1999 (Stats Canada, 1999). Population projections suggest a further growth, in Canada, of 5,600 individuals in the 15-19 age bracket between 2006 and 2011 which is a further growth of 0.3%. This increase in the participation rate has had a potentially negative impact on teaching by increasing the number of students relative to faculty. The student to full time faculty ratio in Canada has increased from 17.5 in 1991 to 22.5 in 2000 and to 24.4 in 2004 (CAUT Almanac, 2004; CAUT Almanac, 2007). At the same time as the participation rate has climbed, the unit cost of educating students has also risen (Chalkley, 1998). The rise in unit cost has been associated with increased uses of technology; increased costs associated with utilities and increased costs of other student services on university campuses.

As a result of increased participation and increased unit costs of higher education, governments are demanding evidence that investments in higher education are being used effectively (Chalkley, 1998). At the same time, publications such as the Maclean's ranking of Canadian universities receive widespread attention from the general public, including current and potential students and their families. Both the public and the government are demanding more frequent and effective evaluation in higher education.

There is an increased focus on undergraduate students acquiring skills that will be useful to them as they enter the workforce (Boyer 1995 p. 6).

> "Many students graduate having accumulated whatever number of courses is required, but still lacking a coherent body of knowledge or any inkling as to how one sort of information might relate to others. And all too often they graduate without knowing how to think logically, write clearly, or speak coherently. The university has given them too little that will be of real value beyond a credential

> that will help them get their first jobs. And with larger and larger numbers of their peers holding the same paper in their hands, even that credential has lost much of its potency."

The emphasis on the student acquisition of skills in their university degrees has fostered an interest in quality assessment. As institutions have increased their focus on skill attainment in response to this demand, there has been an effort to generate evidence that this focus has been successful through the implementation of quality assessments. This focus on student outcomes is highlighted by the widespread participation in the National Survey of Student Engagement (NSSE). The NSSE project was conceived in 1998, piloted in 1999 and more than 1,200 colleges and universities in the United States and Canada have participated since 2000 (NSSE, 2007). NSSE items are based on items that are known to be related to successful college outcomes (NSSE, 2007). One part of quality assessment in NSSE is items associated with assessment of teaching effectiveness which are based on the literature on effective teaching described in Chapter Two.

3.2 Stakeholders in Teaching Evaluation

A number of stakeholders with similar, different or moderately overlapping interests must be involved in the design of a teaching evaluation system (Figure 3.1). The stakeholders can be divided into two groups, those that are internal and those that are external. The internal stakeholders are: students, instructors, instructional staff, departmental chair, faculty (i.e. Dean), Faculty Association, and institution. External stakeholders are: society, parents, employers, provincial and federal governments. The ideal teaching evaluation system must: meet the needs of these various stakeholders;

result in greater student learning; be dynamic to meet the changing needs of the stakeholders and be sufficiently flexible to adapt to different places (i.e., institutions).

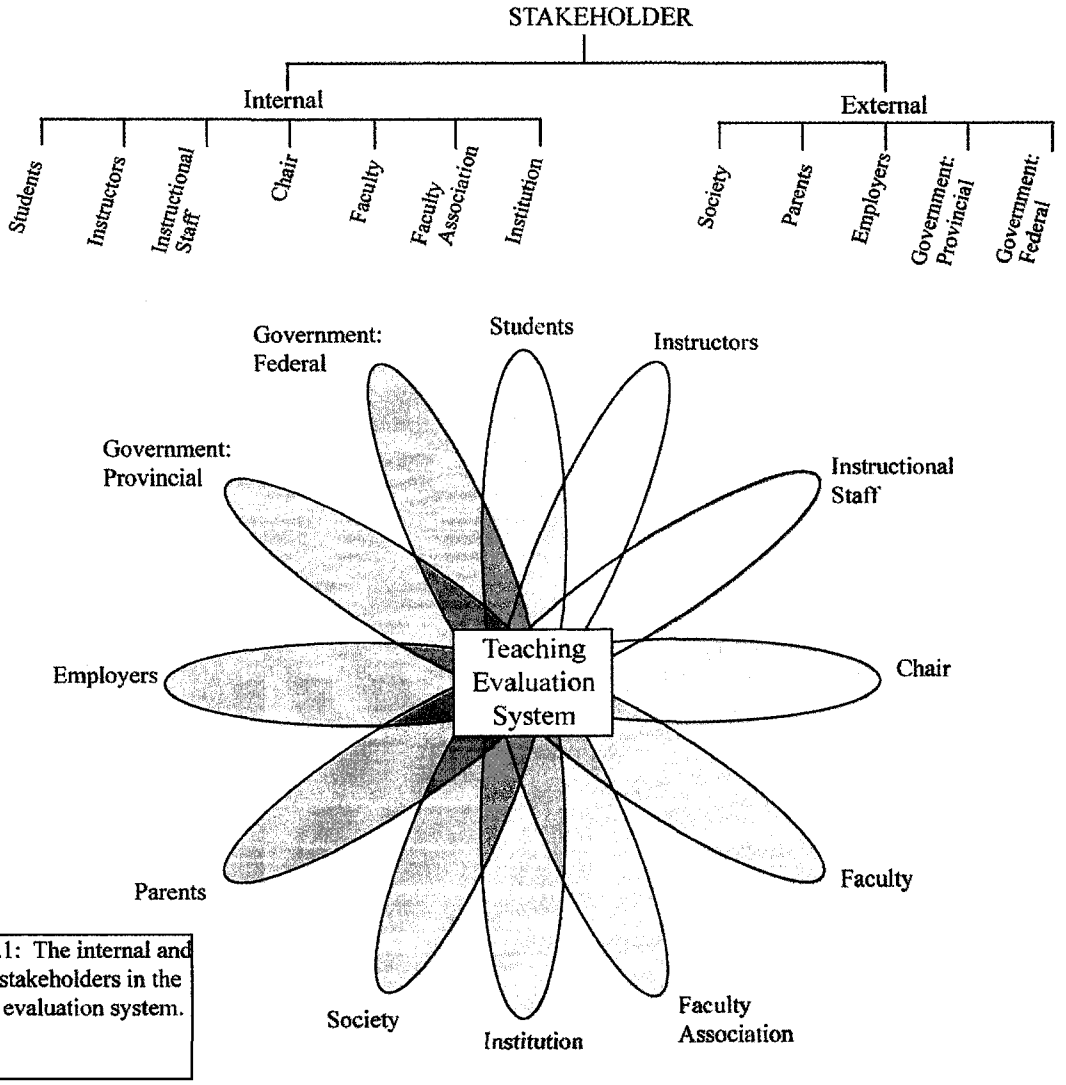


Figure 3.1: The internal and external stakeholders in the teaching evaluation system.

Students require a teaching evaluation system in which they perceive that their voice is being captured and that their comments are being used to help guide enhancements to teaching and learning. Students have a number of reasons for seeking an undergraduate degree. For some it is providing a gateway to subsequent education (e.g., professional school, graduate school). For many students university is seen as a natural progression after high school. Other students may pursue university out of a desire to learn more and a number may pursue their degree because they anticipate it will assist them in securing future well-paid employment. Students are dynamic and in higher education, currently the student population is particularly dynamic with the introduction of Web 2.0 technology (e.g., social networking) (Prensky, 2001). The wide range and evolving expectations for which students pursue post-secondary education combined with the wide range of natural abilities and learning styles results in a diverse range of what students may characterize as effective teaching. As well students are 'new' to the higher education system and may not have a global picture of the breadth of what defines effective teaching in higher education. Thus, students' inability to fully measure teaching effectiveness highlights the need to consider the other stakeholders in the teaching evaluation system.

Instructors also approach university teaching from a wide variety of perspectives. Instructors have a range of experience from fully tenured faculty members to recent appointments on a sessional or contractually limited basis (e.g., 30 years of teaching in higher education vs. teaching a course for the first time). Instructors also have a range of interest in undergraduate education from faculty with a primary interest in undergraduate

education to those with a primary interest in research. All of these individuals will have different demands and levels of interest in the teaching evaluation system.

Instructional staff (e.g., teaching assistants, technicians) may play a substantial role in the delivery of a course and to the delivery of an undergraduate curriculum. This role and contribution is often not effectively evaluated and may benefit from constructive guidance that would be associated with an effective teaching evaluation system.

A ***departmental chair*** may have a variety of interests in the teaching evaluation system. They may be interested in the effectiveness of teaching in order to gauge the impact that teaching is having on departmental budgets (i.e., at some institutions enrolment in courses is directly tied to funding). The departmental chair may also be interested in minimizing the amount of time that they must spend dealing with student complaints and concerns. An effective teaching evaluation scheme may help chairs identify areas of concern to minimize future issues. The chair may have a genuine desire to encourage greater student learning and may perceive an effective teaching evaluation system as a mechanism to contribute to that learning. The chair may be interested in the data gathered in the teaching evaluation scheme in order to use it to assist both program and curriculum development.

The ***Dean*** may have a vested interest in the teaching evaluation scheme for budgetary reasons, complaint management and a genuine desire to encourage effective teaching.

The ***Faculty Association*** or similar group that represents faculty in negotiating a contract between the university and its membership will have an interest in the teaching evaluation scheme. As one chair stated during the interviews, this may not be an interest

in enhancing teaching quality but more of an interest in minimizing culpability of individual members. The Association is genuinely interested in ensuring the reasonable treatment of its members and will likely not have enhancing undergraduate education at the institution as a primary objective. The Association is likely interested in the professional development opportunities for its members. An effective teaching evaluation scheme may provide an opportunity for the Association to provide members a mechanism for non-punitive, professional development.

The ***institution*** is interested in ensuring that its 'brand' (i.e., the profile of place) is highly regarded both in the academic and general community. In order to maintain and enhance its position in these communities, a desire to enhance undergraduate student learning may be sought. Institutions are also in the 'business' of recruiting and retaining undergraduate students. Successful recruitment and retention of students may legitimately be associated with effective teaching which may be enhanced with a good teaching evaluation system. The teaching evaluation system is one mechanism that may be used to ensure accountability in teaching quality which has been shown to be linked to student learning.

Society has an interest in ensuring that a high level of student learning is occurring at universities and that public funds are being appropriately used to benefit student learning. Abler (1994) linked these societal needs for high student learning to the characteristics of an effective geography teacher in higher education when he identified the strong sense of civic responsibility (i.e., community responsibility and outreach) that should characterize effective geography teaching.

Parents usually have a strong interest in their child's undergraduate education. This interest may in-part be driven from financial contributions that are being made to their child's education or from a desire to ensure that their child has an opportunity for a good education. The parent's definition of good teaching is based on a combination of their own experience, their perception of what the experience should entail and professional exposure to teaching and learning. An effective teaching evaluation system should contribute to enhanced student learning. Student learning is typically one part of what parents perceive as a good undergraduate education.

Employers have a desire to have well-educated potential employees graduating from university. The employer's definition of well-educated often places a large emphasis on skill development. An effective teaching evaluation system should help to provide feedback that can be used to enhance the quality of potential employees that are being graduated each year.

In the Canadian context both the ***provincial/territorial*** and ***federal governments*** are stakeholders in higher education. Both governments are interested in ensuring accountability of tax payers' dollars and reputation at the national and international level. This desire for accountability (e.g., maximizing output), combined with a desire for a well-educated population provides the foundation for the interest in an effective teaching evaluation system.

3.3 Goals and Uses of Teaching Effectiveness Assessments

The underlying goal of teaching is student learning; thus, the ultimate measure of teaching effectiveness is enhanced student learning (Marsh, 1987). The ultimate goal for

the assessment of teaching effectiveness is to enhance teaching quality which in turn should lead to increased student learning. Direct measurement of student learning gains is challenging for a number of reasons. Often student learning is evaluated based on a grade received in a course, but this measure does not consider the starting point (i.e., what was the level of knowledge, skill and understanding at the start of the course). Ideally, to measure student learning gains, extensive pre- and post-testing of students' knowledge, skills and understanding should occur. However, this rarely occurs. Learning in a subject does not begin and end with the start and end of a course; thus, longitudinal studies of students' retention and/or subsequent gain of knowledge, skills and understanding as a result of a course should be measured. Again, this rarely occurs; thus, more pragmatic goals for teaching effectiveness assessments are often identified. The five most widely acknowledged purposes of teaching effectiveness assessments are:

- to provide diagnostic feedback to faculty about their teaching, so that faculty can use this information to enhance their teaching (Murray, 1980; Marsh, 1987; Marsh and Roche, 1997);
- as a measure to be used for administrative purposes to assist in guiding their decisions about promotion, tenure and salary(Murray, 1980; Marsh, 1987; Marsh and Roche, 1997);
- to provide information to prospective students to assist them in their selection of courses and instructors (Murray, 1980; Marsh, 1987);
- to assess the quality of individual courses to be used for course and curriculum improvement and design (Marsh, 1987), and,
- to provide data for research on teaching (Murray, 1980; Marsh, 1987).

There is disconnect between the underlying goal of teaching (i.e., student learning) and the purposes widely acknowledged for completing teaching effectiveness assessments. At some level, each of the above goals may be attributed to enhanced student learning,

but the outcome from the assessment may or may not be used to enhance student learning. For example, the data from teaching effectiveness assessments may be used to conduct research, but will only enhance student learning if the outcomes from that research are translated back into the student learning environment.

3.4 Elements of Effective Teaching Assessments

This section will identify the elements (i.e., characteristics) necessary for a teaching effectiveness assessment, focusing primarily on student evaluation of teaching effectiveness (SETE). Each element will be described and its relevance to SETEs will be discussed. Even twenty years ago a voluminous literature of more than 1,300 articles and books described the evaluation of teaching via SETEs (Cashin, 1988). A Google Scholar search in 2007 identifies 216,000 items about student evaluation of teaching effectiveness.

a. Identify assessment goals

The first and foremost element in designing a good assessment of teaching effectiveness is the identification of the goal of the assessment exercise (Diamond and Adam; 1993; Braskamp and Ory, 1994; Cashin, 1996). The possible goals include: to provide diagnostic feedback, to provide a measure for administrative purposes, to provide information to prospective students, to provide input to course and curriculum design and to provide data for research on teaching (see Section 3.1). In some cases, a single assessment may be designed that will satisfy more than one of these goals, but a single assessment is unlikely to satisfy all of the goals (Derry, 1979; Murray, 1980). There is widespread agreement that when assessing ***students*** that the goals (i.e., criteria) must be

evident, published and discussed with the students (Woolf, 2004). Similarly, when assessing ***faculty,*** the goals must be well-defined, clearly articulated and well understood to ensure that an effective, meaningful evaluation occurs (Wright and O'Neil, 1992). SETE are frequently used as an end-of-course assessment tool to provide input for administrative decisions and to provide instructor feedback. These data are also sometimes used in the preparation of 'anti-calendars' used by students to help them select courses (Murray, 1980; Marsh, 1987).

b. Determine type of assessment required – formative vs. summative

Once the goals for evaluating teaching have been clearly identified, then it must be determined if the goals are conducive to a formative or summative evaluation (Cashin, 1996). In general, a single assessment is unlikely to provide useful formative and summative feedback (Derry, 1979; Murray, 1980; Cashin, 1989). A formative evaluation is used to satisfy evaluation goals that require feedback to shape and guide an ongoing activity. An example of a formative evaluation would be a mid-year teaching evaluation in which an instructor asks students for feedback about the introduction of a new teaching technique and then uses the student feedback to adjust their use of the technique for the remainder of the course. A summative evaluation is used to respond to a goal that is final or terminal. An example of a summative evaluation goal would be determining if Candidate A has demonstrated sufficient teaching effectiveness to receive a promotion. SETE are most commonly summative evaluations provided at the end of a course with the primary aim of providing input for administrative decisions (Marsh, 1987). They also provide some feedback to the instructors on teaching effectiveness, particularly if they include a section for students' written comments (Marsh, 1987).

c. Ensure credibility of assessment tool

Braskamp et al. (1984) argued that there must be a perceived trust between the participants in a teaching evaluation scheme, much like the practice of participatory management in business. Teaching evaluation participants include: faculty, administrators, students, alumni and more recently, the public. Ensuring participant participation in the design and implementation of the evaluation scheme enhances credibility (Wright and O'Neil, 1992). Murray (1973) stated that student feedback about teaching effectiveness will not be taken seriously until there are more tangible rewards for teaching improvement. He argued that students would not put the necessary effort or approach the task seriously until they perceive an impact from their involvement. Theall and Franklin (2001) argued that the poor opinion that many faculty have of SETE is their lack of knowledge regarding the course evaluation literature and the research rigour that has been applied to understanding potential bias in SETE (see validity discussion in Section *l* below). LaCelle-Peterson and Finkelstein (1993) emphasized the importance of having the support of senior faculty to the mission of enhancing teaching and effective evaluation. Senior faculty are the mentors of junior faculty; thus, they carry considerable weight in guiding junior faculty (LaCelle-Peterson and Finkelstein, 1993). The participation of campus leaders, including student leaders, faculty leaders and administrative leaders, will provide increased levels of acceptance and ownership of the evaluation scheme (Cashin, 1996).

d. Ensure a broad evaluation system

A good evaluation system must measure teaching effectiveness beyond what happens in the classroom and incorporate a broad definition of teaching (Braskamp et al.,

1984). In order to evaluate what occurs beyond the classroom, students cannot be used as the sole data source (Wright and O'Neil, 1992; Murray et al., 1982). Using multiple sources of data, including students, peers and trained evaluators (i.e., beyond only using SETE which consider students as the sole data source) can help ensure a broad evaluation scheme is being used (Cashin, 1996).

e. Provide consultation/feedback

The greatest benefit from teaching effectiveness evaluation occurs when a consultation process is included as part of the feedback process (Wilson, 1986; Cashin, 1996). Feedback should be constructive (Cashin, 1996). The value of this feedback process in enhancing instructors' abilities to articulate their strengths and weaknesses has been demonstrated in the work of Roche and Marsh (2000). Roche and Marsh (2000) examined the level of agreement between the self-concept of instructors (i.e., their own perceptions of their teaching effectiveness) who had not previously received SETE results to their SETE results and the self-concept of instructors, who had previously received SETE results, to their SETE results. The outcome demonstrated a modest (median r=0.20) level of agreement for teachers who had not previously received feedback and a substantial (median r=0.40) level of agreement for teachers who had previously received feedback. Thus, SETE had the desired outcome of increasing aninstructor's awareness of the areas students perceived as the instructor's strengths and weaknesses. There is widespread agreement across many fields that self-concept beliefs can have important influences on motivation and behaviour of individuals (Marsh, 1990; Hattie, 1992; Marsh and Craven, 1997).

f. Provide a reward system

Cashin (1996) argued that for an effective teaching evaluation system to exist (i.e., one that changes faculty behaviour in such a way as to enhance teaching), it must effectively discriminate between teaching effectiveness of different faculty members, faculty members must perceive these discriminations as accurate and faculty members must be treated based on these discriminations (i.e., be rewarded differently). Wright (1999) supported the argument that a reward system needs to recognize excellence and improvement in teaching and not simply recognize adequate teaching. The system should be a supportive system, not a punitive system (Wright, 1999). Abler (1994) reported that 70% of US institutions had reward systems in place by the early 1990s although no evidence is provided about the effectiveness of the reward systems.

g. Ensure a supportive culture exists

The culture in which teaching and learning occurs contributes to faculty motivation towards teaching and to the level of student learning. Ramsden (1979) surveyed 285 students in six departments (social science, applied science, natural science, two arts departments and independent studies) at Lancaster University and followed the paper survey with semi-structured interviews with a minimum of ten students from each department. The outcome of this work was that the learning environment was shown to be very important to students' perceptions of their learning (Ramsden, 1979).

Feldman and Paulson (1999) reviewed the existing research literature consisting of qualitative studies, case studies and surveys on teaching culture. They identified the following seven characteristics of a supportive teaching culture (Feldman and Paulson, 1999):

- high level of administrative commitment and support;
- faculty involvement, shared values and ownership;
- broad definition of scholarship;
- teaching demonstration or pedagogical colloquium as part of the hiring process;
- frequent interaction, collaboration and community between faculty and teaching and learning centres;
- supportive and effective chairs; and,
- connection of rigorous evaluation of teaching to tenure and promotion decisions.

One weakness of many SETE schemes is instructors' perception that the instruments are not credible often due to the perception that the instruments are not rigorous (e.g., valid, reliable). This lack of credibility creates a culture of distrust and can negatively impact the success of a teaching evaluation scheme.

h. Multi-dimensionality

Assessments of teaching effectiveness must recognize that teaching is a multi-dimensional activity (McKeachie, 1997; Marsh and Roche 1997). A large number of factor analytic studies have demonstrated that SETE are multi-dimensional (e.g., Feldman, 1976; Cohen, 1981; Feldman, 1989). A multi-dimensional SETE means that the tool is measuring several aspects of teaching and that no single rating element or set of rating elements will be useful for global evaluation purposes. Several meta-analytical studies have been repeatedly cited for their identification of these dimensions which can be found in Table 2.1 (Marsh, 1984; Feldman, 1984; Feldman, 1988). The multi-dimensional aspect of SETE must be considered when interpreting SETE results and it is not appropriate to average all the items (Cashin, 1988).

i. Reliability

Reliability is most commonly determined from item analysis results (i.e., high correlations between responses to different items that are designed to measure the same dimensions) and from studies of inter-rater agreement (i.e., among ratings by different students in the same course). Effective teaching assessments must fulfill both reliability criteria. Internal consistency between responses to different items that are designed to measure the same dimensions are consistently high (Marsh, 1987). SETE have high inter-rater agreement provided, class averages of items in medium- to large-sized classes are being evaluated. A frequently cited SETE is the SEEQ (Students' Evaluations of Educational Quality Instrument developed by Marsh, 1987). The SEEQ has an inter-rater agreement between two students in the same class typically in the 0.20s which is low-level agreement; however, the reliability for SEEQ factors increases dramatically when more raters are compared. The correlation is approximately 0.74 from 10 students, 0.90 from 25 students and 0.95 from 50 students (Marsh, 1987). This confirms the reliability of a well-designed SETE.

j. Stability

An effective teaching assessment tool must ensure stability of responses by individual raters over time (i.e., an effective tool must ensure continuity in an individual's evaluation of an instructor over time, often considered to be at the time of course completion and several years hence). A common argument raised to dismiss SETE is that students are perceived to be incapable of completing effective teaching evaluations at the time they are enrolled in a course and that they are only able to effectively evaluate

once they can reflect on their learning experience after applying it in later coursework or after graduation. A longitudinal study (Marsh and Overall, 1979; Overall and Marsh, 1980) compared the end-of-course SETE with ratings by the same students years later (minimum of one year after graduation) and determined an average correlation of 0.83. This confirms that SETE results from individual raters are temporally stable.

k. Generalizability

The effective teaching assessment tool will measure the instructors' overall effectiveness, not just their effectiveness within a specific course in a single term. Marsh (1984) analysed data from 1,364 courses and divided the results into four categories: same instructor teaching the same course, same instructor teaching a different course, different instructor teaching the same course and different instructor teaching a different course. This division allowed the researcher to consider the effect of instructor and the effect of course on the SETE. Items on the SETE were separated into those related to the instructor (e.g., enthusiasm, organization, group interaction, individual rapport) from those more related to the course (e.g., learning/value, workload/difficulty). Marsh (1984) successfully demonstrated, that it is the instructor and not the course which is the primary determinant of the student ratings (Table 3.1).

Table 3.1: Average correlations of 1364 courses analysed to determine the impact of course and instructor on ratings of SETE (Marsh, 1984).

Measure	Same Teacher Same Course	Same Teacher Diff Course	Diff Teacher Same Course	Diff Teacher Diff Course
Student Rating				
Learning/Value	.696	.563	.232	.069
Enthusiasm	.734	.613	.011	.028
Organization/Clarity	.676	.540	-.023	-.063
Group Interaction	.699	.540	.291	.224
Individual Rapport	.726	.542	.180	.146
Breadth of Coverage	.727	.481	.117	.067
Examinations/Grading	.633	.512	.066	-.004
Assignments	.681	.428	.332	.112
Workload/Difficulty	.733	.400	.392	.215
Overall Course	.712	.591	-.011	-.065
Overall Instructor	.719	.607	-.051	-.059
Mean Coefficient	.707	.525	.140	.061
Background Characteristics				
Prior Subject Interest	.635	.312	.563	.209
Reason for taking (% indicating general interest)	.770	.448	.671	.383
Class Average Expected Grade	.709	.405	.483	.356
Workload/difficulty	.773	.400	.392	.215
Course Enrollment	.846	.312	.593	.058
% attendance on day evaluations administered	.406	.164	.214	.045
Mean Coefficient	.590	.540	.491	.211

Generalizability of the results is particularly important if the SETE results are being used to make personnel decisions (e.g., tenure, promotion) based on the individual's effectiveness as a teacher (Cashin, 1988).

l. Validity

Validity is a term describing how well a measure accurately reflects the concept it was intended to measure (Babbie, 2001). Construct validity of SETE attempts to demonstrate that student ratings are logically related to various other indicators of effective teaching and less correlated with other variables (Marsh and Roche, 1997).

Evidence for construct validity support comes from the long-term stability and generalizability of SETE (Marsh, 1987).

Construct validity of SETE is supported through assessment of student learning because, in theory, the students of more effective teachers will learn more. Both Cohen (1981) and Feldman (1989) completed meta-analysis of student teaching evaluations from multiple-section courses where different instructors taught different sections of the same course using the same syllabus, textbook and final exam. The average correlations between instructor effectiveness measured on the course evaluation and student achievement or learning measured by student grade on a common final exam is 0.47 and 0.46 respectively (Cohen, 1981; and Feldman, 1989). These types of studies are somewhat problematic because they can only be supported with evidence from large, multi-section courses and the methods can not be applied in smaller classes.

A third source of support for construct validity of SETE is comparison of instructor self-ratings with SETE. Ten studies that examined instructor self-ratings with student ratings had correlations of 0.20 to 0.69 with an average of 0.41 (Marsh, 1984).

A fourth source of support for construct validity of SETE is a comparison of peer ratings and SETE ratings. Early work by Kulik and McKeachie (1975) demonstrated correlations of 0.47 to 0.62 when student ratings are correlated to administrator ratings. Kulik and McKeachie (1975) found student ratings correlations with colleagues' ratings of 0.48 to 0.69. Numerous other studies have been unable to substantiate these findings (e.g., Centra, 1979; Braskamp et al., 1984).

The validity of teaching effectiveness assessment tools is brought into question when potential biases are identified. Bias includes all the variables that may affect

ratings on SETE that are not a function of instructor effectiveness (Cashin, 1988). A large number of variables that have been identified as potential biases, but show little or no impact to student ratings, include:

- student's cultural background (Watkins and Akande, 1992);
- instructor's age and teaching experience (Feldman, 1983);
- instructor's research productivity (Feldman, 1987);
- student's age (Menges, 1973);
- student's academic standing (e.g., GPA) (Feldman, 1976);
- time of day course is taught (Feldman, 1978); and,
- class size, smaller classes tend to have slightly higher evaluations but the relationship is quite weak with an average r = -0.09 (Feldman, 1984; Marsh, 1987).

A large number of variables have been identified as potential biases and show a relationship to student ratings (Table 3.2).

Table 3.2: Identification of potential biases to SETE, an explanation of the bias and in some cases an explanation that the perceived bias may in fact not be a bias

Potential Bias	Description	Explanation
Student motivation	Students tend to give higher ratings to courses in which they had prior interest (Marsh, 1984) or to courses in which they are taking as an elective (Feldman, 1978; Marsh, 1987).	
Course level	Higher level courses, particularly graduate level courses tend to receive higher evaluations (Braskamp et al., 1984).	
Anonymity	Students tend to provide higher ratings when they must identify themselves (Feldman, 1979; Braskamp et al., 1984; Marsh, 1984).	
Identified purpose of the rating	Students tend to provide higher evaluations if they are informed that the ratings will be used for personnel decisions as opposed to being used strictly for instructor self-improvement (Feldman, 1979; Braskamp et al., 1984; Marsh, 1984; Marsh, 1987).	
Anticipated grades	There is a small positive correlation between student ratings and anticipated grades i.e., r = 0.10 to 0.30 (Feldman, 1976; Marsh 1984; Marsh, 1987)	Three hypothesis have been provided to explain this correlation: • the student anticipated a high grade, the instructor gave higher grades than the students earned and correspondingly the students gave the instructor higher ratings than earned; thus, the student ratings are not valid (Cashin, 1988). • the students who anticipated receiving higher grades gave the higher ratings and they were the students that had learned more; thus, the student ratings are valid (Cashin, 1988). • the higher ratings are being controlled by a student characteristic (e.g., motivation) which in turn results in the student learning more, thus, earning a higher grade and giving a higher rating making the ratings valid (Cashin, 1988).

Potential Bias	Description	Explanation
Academic discipline	Some fields of study tend to receive higher ratings (Feldman, 1978; Braskamp et al., 1984; Marsh, 1984).	The higher ratings tend to occur in social science fields whereas the lower ratings tend to occur in the math and science fields. If the lower/higher ratings are a function of the discipline then there is a bias; however, if the ratings are a function of better-quality instruction and greater learning in some disciplines than others, it is not a bias (Cashin, 1998).
Sex of the instructor & sex of the student	Female instructors tend to receive slightly lower evaluations from male students than from female students; whereas, male instructors tend to receive similar evaluations from students regardless of the sex of the student (Basow, 1994; Basow and Distenfeld, 1985)	Theall and Franklin (2001) identify that the female instructors in the department studied were disproportionately assigned large entry-level required courses while their male colleagues taught disproportionately upper-level and graduate seminars; thus, it may not be a bias of the rating but rather a function of the courses that instructors are assigned may not be gender neutral. Other studies that have considered gender have demonstrated no strong or regular pattern of gender-bias in SETE (Feldman, 1992; Theall and Franklin, 2001).
Course difficulty & workload	Students tend to evaluate courses in which they have to work harder a higher rating (Marsh, 1984; Cashin, 1988).	If this work is an indicator of greater student learning then it is not a bias. As well, this bias is in the opposite direction to what would be anticipated if it was an indicator of bias. This is further evidence that it is unlikely to be a bias (Marsh, 1987).

According to Marsh (1987) a number of methodological weaknesses are common to many of the studies that identify a bias relationship in student ratings including:

- using correlation to argue for causation. A strong correlation simply shows that a concomitant relationship exists;
- neglecting to consider the distinction between practical and statistical significance;
- failing to consider the multivariate nature of student ratings and a set of potential bias;
- selecting inappropriate units of analysis. SETE are nearly always considering class average results; thus, the appropriate unit of analysis is the class. Often bias is identified using individuals students as the unit of analysis and the same bias is not demonstrated when the class is considered as a whole;
- failing to examine replicability of findings in similar settings and the inability to generalize the results to other settings;
- lacking an explicit definition of bias. If a factor impacts teaching effectiveness and this impact is measured in the SETE then it is not bias; and,
- selecting appropriate experimental manipulations. Experimental manipulations must ensure the validity of the manipulation and the representativeness of the experiment.

The above discussion demonstrates that, although there are a wide number of potential biases to SETE validity, many have reasonable explanations (Table 3.2) and only a few are likely to impact the data.

The preceding sub-sections (*a-l*) have demonstrated that there is a wide variety of elements which comprise an effective teaching evaluation and that SETE is a reasonable tool to address many of these elements. The following section will describe alternative methods to measure teaching effectiveness.

3.5 Tools to Measure Teaching Effectiveness

Course evaluations (SETE) are not the best tool to measure teaching effectiveness because course evaluations rarely assess the wide breadth of what is involved in teaching. They generally only measure what occurs in the classroom (Cashin, 1995). There is a wide range of alternative assessment tools available for evaluating teaching. Centra (1977) surveyed 453 chairs of departments to rate the current use and importance of 15 different kinds of data used to evaluate teaching (Table 3.3).

Table 3.3: Ranking of current use and importance and preferred use and importance of sources of information for evaluating teaching effectiveness according to 453 chairs of departments. Departments are subdivided into Research Universities (R), Doctoral Granting Universities (D); and Comprehensive Universities & Colleges (C) according the to the Carnegie Institutional Classification of 1973.

	Current Use and Importance				Importance Each Should Have			
Source of Information	All n=453	R n=158	D n=122	C n=173	All n=453	R n=158	D n=122	C n=173
Chairman evaluation	1	3	1	1	2	2	2	1
Colleagues' opinion (tied)	2.5	1	3	2	3	3	3	4
Systematic student ratings (tied)	2.5	2	2	4	1	1	1	2
Committee evaluation	4	4	4	3	4	4	4	3
Informal student opinion	5	5	5	6	6	6	7	7.5
Dean evaluation	6	8	6	5	8	11	9	6
Content of course syllabi and examination	7	7	8	7	5	5	5	5
Popularity of electives	8	6	7	10	13	10	14	13
Self evaluation or report	9	10	9	8.5	11	14	10	10
Teaching improvement activities	10	11	11	8.5	10	12.5	11	7.5
Student examination performance	11	9	12	11	12	9	13	12
Colleague ratings based on classroom visits	12	12	10	12	7	7	6	9
Alumni opinions or ratings	13	13	13	14	14	12.5	12	14
Long-term follow up of how students' perform	14	14	14	13	9	8	8	11
Videotape of classroom teaching	15	15	15	15	15	15	15	15

This list has been used by other researchers including Seldin (1980 and 1984), to identify potential data sources for information regarding teaching effectiveness. Table 3.3 demonstrates general agreement between the perceived 'current use and importance' and the perceived projected use and importance of different evaluation tools that are ranked in the top three. The amount of variation between actual and projected use of evaluation tools increases as their importance decreases. Cashin (1989) identified a critical weakness in Table 3.3 that four of the top six data sources listed above are not actually data sources but are evaluators (i.e., chairman evaluation, colleague opinion, committee evaluation, and dean evaluation). These evaluators are likely using the data that appear elsewhere on this list to inform their opinions. It is interesting to note in Table 3.3 that there is near universal agreement that systematic student ratings (SETE) should receive greater importance than currently and that the items identified by Cashin (1989) should all receive less importance than they currently receive.

The following sub-sections (*a-e*) will review a number of potential teaching assessment tools.

a. Self evaluation

Self-reflection is often seen as a useful formative method of evaluation provided that the evaluation is descriptive rather than evaluative (Cashin, 1989; Chalkley et al., 2000). Centra (1993) surveyed 343 instructors and their students from five US colleges to investigate the relationship between student evaluation of instructors and instructors' self-reports. The outcome of this work was a general disagreement between student evaluations and self-ratings (median correlation of 0.21 on the 21 items). Instructors

generally gave themselves better ratings than their students, although between five to six percent of instructors gave themselves much lower. The greatest discrepancies were in items related to student-teacher interaction, clarity of course objectives and the instructor's openness to other viewpoints. Self-report may be an important tool when combined with SETE to point out discrepancies between student perspectives and instructor perspective (Centra, 1993).

b. Student evaluation

SETE are a widely used method of collecting student feedback on instruction and have been previously discussed in Section 3.3. A number of other models to collect student feedback are described by Cox (2000) including:

- using students as recorders/observers in the classroom;
- using a 'faux' student in the class;
- having a student videotape a lecture;
- interviewing students within the class as part of a focus group; and,
- using a student consultant trained as a classroom observer.

Alumni surveys of overall teacher effectiveness have demonstrated a fairly high correspondence with SETE. (Druckers and Remers, 1951; Marsh and Overall, 1979; Overall and Marsh, 1980).

Delayed measures of retention, which attempt to measure the degree of retention of subject material months to years after a course has ended, can be used as a measure of teaching effectiveness (McKeachie, 1958) The basis of this argument is that students who demonstrate a high level of retention after a course has ended are demonstrating higher levels of subject mastery rather than just surface learning of subject content which may occur when only final exams/grades are used to measure student learning. Although

the logic to this argument is relatively sound, it is extremely difficult to measure because of the large impact of confounders such as other related courses that a student may have taken, other instructors that the student may have experienced and the applicability of the subject material in the student's eventual career.

Assessing teaching effectiveness by determining the number of students a first-year instructor is able to recruit into a departmental program following their completion of a first-year course has some validity (McKeachie, 1958). This is particularly true in the 'geo' field because many program students only become interested in the discipline after taking a first-year course and learning about the nature of the discipline (O'Connell et al., 2003). One downfall of this measure of teaching effectiveness is that it may promote interdisciplinary competition as opposed to a common concern for the best possible education.

c. Peer/Colleague Evaluation

Peer evaluation of teaching is irregularly used in the US, except for decisions regarding promotion and tenure (Chalkley et al. 2000). There is often some confusion in the literature and among practitioners about what is meant by peer evaluation. For the purpose of this work, peer observation will refer to the practice where a peer observes a colleague teaching on one or two occasions. Peer evaluation involves a variety of tools to inform the evaluation which may include peer observation, review of student assessment tools (e.g., tests, and assignments), course outlines, and student focus groups. Peer review of teaching is promoted by Gibbs (1999) as a method to create a robust, reliable method of evaluating teaching. Peer evaluation is considered an important aspect to increase the professionalism and scholarship of teaching. Rice (1986) argued that peer

evaluation has long been considered one of the key characteristics of a profession and is evident in both medical and legal professions. As well, peer evaluation is the foundation on which research excellence is measured. According to Murray (1980), peer evaluations when compared to SETE are less sensitive, less reliable, less valid, more threatening and disruptive of faculty morale and more affected by non-instructional factors such as research productivity. This strong criticism of peer evaluation may be the result of the lack of rigorous procedures that peer evaluations, of teaching effectiveness, have historically followed and that this type of evaluation is often perceived to be based on opinion rather than fact (i.e., it is subjective rather than objective).

d. Evidence of methods that contribute to effective learning

Angelo (1996 p. 59) stated that an approach to assessing teaching effectiveness is to look for teaching methods that enhance student learning.

> "Recent research in psychology, cognitive science, and education has demonstrated that certain conditions, practices, and processes are more highly correlated with learning than others. So, by looking for and assessing indicators that are strongly correlated with effective learning, we can improve our assessment and evaluation of teaching."

Increased student engagement and deep learning are two approaches that have been documented to enhance student learning (Ramsden, 1992).

Research has demonstrated that much of student learning occurs outside the classroom and is guided by the assessment tasks (e.g., lab assignments, essays, tests) set by the instructor (Snyder, 1971; Pascarella and Terenzini, 1991; Knapper and Rogers, 1994). The importance of assessment tasks for student learning makes the quality of the assessment tasks an important contributor to the effectiveness of the instructor and an element that should be considered.

e. Teaching Dossier

Teaching must be documented (Shulman, 1993). One form of documentation is the teaching dossier (sometimes called portfolio) which is a factual description of an instructor's major strengths and accomplishments in teaching. It provides a forum to document the breadth and quality of teaching and has been compared to the lists of publications, grants and honours used to describe research excellence (Seldin, 1991). Seldin (1991) provided a comprehensive list of the possible components of a teaching portfolio. These are:

- statement of teaching responsibilities;
- a reflective statement of the instructor's teaching philosophy, strategies and objectives;
- a personal statement of the instructor's five-year goals;
- representative course materials including syllabi, assignments and examinations and explanations of why they were constructed in their current form;
- descriptions of activities to evaluate and improve one's teaching;
- description of participation in curricular activities;
- self-reflection of teaching abilities;
- contributions to scholarship of teaching; and,
- summary of activities related to the supervision of undergraduate theses.

Teaching portfolios are being increasingly used in the US in a desire to add scholarly rigour to teaching evaluation (Abler, 1994; Edgerton et al., 1991). Smith (1991 p. 65) as the Commissioner of the *Commission of Inquiry on Canadian University Education* identified as an action item:

> "teaching dossiers should be widely adopted as a basis for evaluating the teaching record of faculty."

An effective teaching evaluation scheme should include a variety of methods to assess effective teaching. The following sections will consider the opportunities and constraints to the development of teaching assessment tools in geography.

3.6 Opportunities for Development and Application of Teaching Assessment Tools in Geography

The Association of American Geographers in a Special Committee Report, *Toward a Reconsideration of Faculty Roles and Rewards in Geography* compiled by R. Abler (1994 p. 14) identified as the number one recommendation that:

> "competent teaching – verified by rigorous peer review – be a *necessary condition* of retention and advancement in all professorial positions in geography in all academic institutions."

The identification and emphasis placed on 'rigorous peer review of teaching' by one of the largest professional geography associations in the world, provides an enormous impetus for the development and application of effective teaching assessment tools in geography.

Instruction in effective teaching should be provided for Ph.D. students, particularly for candidates who intend to pursue academic careers (Abler, 1994). Junior faculty who have been instructed in effective teaching are likely to be more receptive to the development and the application of effective teaching assessment tools. Geographers are viewed by those outside the discipline in higher education as being in the 'vanguard' of developments and commitment to the professionalism of teaching (Brown et al, 2002) and

> "geography has a strong reputation for distinctiveness, effectiveness and originality of its teaching" (Cooke, 1998 p. 238).

The staff of geography departments, when surveyed as part of accountability studies in the UK, were found to be enthusiastic and committed to teaching (Chalkley, 1998). To maintain and enhance this role as leaders in the trend to professionalize teaching in higher education, it is reasonable to assume that geographers will view the development and application of effective teaching assessment tools in a favourable light.

Teaching is perceived as a critical role in geography programs because few students enter university with the intent to major in geography, rather they are 'enticed' to enter the field after a positive experience in a university-level geography course (O'Connell, 2003; Diamond and Adam, 1995; Abler, 1994). The high value placed on teaching by the discipline creates an environment conducive to evaluation as a means to enhance teaching effectiveness.

Geographers have a strong desire to maintain and enhance the geographic literacy of the population. This desire for geography awareness and knowledge transfer provides impetus to ensure the greatest amount of student learning may occur through effective teaching.

The integrative and synthetic nature of geography requires instructors who present information in a clear and coherent fashion (Diamond and Adam, 1995; Abler, 1994). The requirements placed upon the instructor as a result of the nature of the discipline makes teaching a particularly important role within the field of geography (Diamond and Adam, 1995; Abler, 1994). This important role of teaching effectiveness in geography increases the emphasis that geographers will place upon the need for the development and application of teaching evaluations.

Geography involves a wide breadth of skills and content. This breadth provides geography instructors with the opportunity to use a wide variety of teaching and learning techniques. Geography students 'overwhelmingly' express satisfaction with the quality of their teaching (Chalkley, 1998). The geography instructor is constantly trying to maintain a balance between knowledge transfer and skills acquisition when teaching. As a result of trying to maintain this balance, it is reasonable to assume that geography instructors are genuinely interested in their ability to maintain this balance and ensure student satisfaction levels are maintained. This interest lends itself to the development and application of teaching evaluation schemes within the field of geography.

Geographers have a culture of critical thinking and self reflection. This culture was evident at the recent Canadian Association of Geographers Annual Meeting 2004 in Moncton at which a session entitled 'Can GIS Save Geography?' was held. This reflective culture is not a recent development in geography since previous CAG meetings have hosted similar reflective sessions such as the session at the 1995 meeting, 'Is Geography Sustainable?' (Nelson and Semple, 2000). The reflective culture in geography creates an environment where reflection and assessment are perceived as a positive and important component of any task. Logically this would include the development and application of teaching assessments.

The importance placed on geography education in high school curricula is declining. Within some provinces there are no mandated geography courses as part of the high school curriculum. Universities, particularly those within the province of Ontario, have experienced a decline in geography program enrolment numbers in recent years. Effective teaching may be one way to halt the decline in student participation in

geography. If teaching is perceived by instructors as important to the 'survival' of their discipline, this would enhance their desire for the development and application of teaching evaluation tools within geography.

Historically, a large number of students who have completed geography degrees have entered the field of teaching. Instructors are aware that many of their students will enter the field of teaching and this combined with faculty members' desire to enhance geography awareness creates a culture in which teaching is valued and tools to effectively measure teaching quality would be sought.

3.7 Constraints for Development and Application of Teaching Assessment Tools in Geography

The creation of a single assessment in the discipline of geography is challenging because the discipline is extremely broad, encompassing both physical and social sciences. This breadth is also evident in the wide range of environments in which geography teaching occurs (e.g., small to large lectures, seminars, laboratory, and field work). This breadth of subject material and teaching environments impedes the development and application of a single effective teaching assessment tool.

Geography teaching involves a large component of skills-based instruction. As stated earlier, the ultimate measure of effective teaching is enhanced learning. Skill development is very difficult to measure. Since skill acquisition is a large component of student learning in geography, it is imperative that measurement of teaching effectiveness considers the level of skill development and growth that occurs in a course. The

challenges involved in teaching skills and evaluating skills would constrain the development of effective teaching evaluation tools.

Field teaching is an integral component of geography teaching and can occur as a short, several-hour experience or may exist as an intensive one-to two-week field camp. Many of the questions contained within a typical SETE are not relevant to field teaching. There are many positive outcomes from these field camps that would be challenging to measure during or immediately following the field course. Students gain tremendous confidence in their abilities to apply their knowledge and skills in a field setting. A well taught field camp will also lead to a valuable increase in student involvement within the department and increased student-to-student interaction. Student involvement has been demonstrated to be positively related to student engagement which has been shown to enhance student learning. The learning benefit from a field course exists beyond the course itself. Measuring this type of teaching effectiveness is very challenging and would constrain the development of teaching assessment tools.

Geography departments tend to be modest-sized academic units on the university campus. In general, any type of accountability review is more challenging in smaller departments (Smith, 1988). Thus, the relative size of geography departments will constrain the development and application of teaching effectiveness assessments. The availability of government funding for education impacts the activities, including teaching-related activities that can occur. Recently, funding for higher education within Canada has seen a substantial decline (Robinson, 2001). This decline in available funds may constrain the development and application of teaching evaluations. The examination of teaching effectiveness within geography and the discussion of teaching assessment

tools provide the foundation for the development of a research project that will examine in more detail teaching effectiveness assessments within Canadian geography departments.

3.8 Conceptual Framework – Teaching Evaluation

The thorough review of the literature presented in this chapter has permitted the researcher to develop a conceptual framework to describe teaching evaluation practices (Figure 3.2). This conceptual framework illustrates that teaching evaluation practices begin with a demand which may originate with one or more of the stakeholders. From the demand for an evaluation, a goal is identified and an evaluation process is implemented. The evaluation process begins by identifying the type of assessment required (i.e., formative or summative). Then the appropriate tool is selected ensuring a number of criteria are met including: credibility, multi-dimensionality, reliability, stability, generalizability and validity. This process all occurs within an environment with a supportive culture, a rewards and incentive mechanism and a mechanism for feedback. The conceptual framework that has been developed through the review of the literature will be examined through interviews with geography chairs, university administrators, staff in teaching and learning centres and undergraduate students in order to test it.

Chapter 4 will describe a research methodology that will inventory current definitions of good teaching and teaching evaluation practices within Canadian university geography departments and compare those practices with the two conceptual frameworks that have been developed to describe effective teaching and a teaching evaluation system.

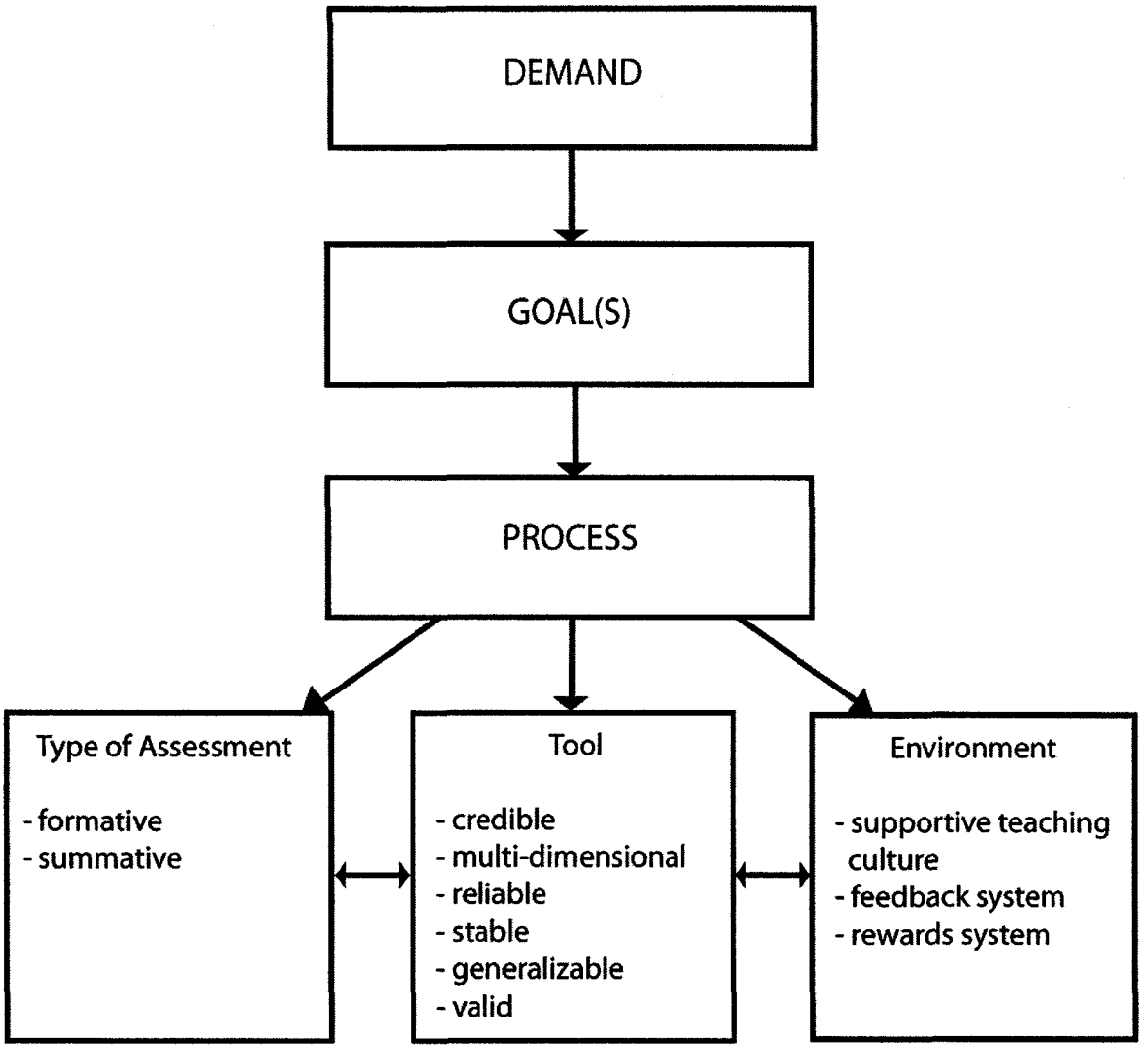


Figure 3.2: A conceptual framework for teaching evaluation practices.

CHAPTER FOUR:

Research Design

4.1 Introduction

The review of the literature (Chapter Two and Three) identified and described the overall characteristics of effective teaching in higher education and those characteristics that are specific to geography. The literature review also considered different methods for evaluating teaching. This literature was then used to create two conceptual frameworks, one to describe effective teaching and one to describe teaching evaluation practices. This knowledge has been used to develop a methodology and series of associated research instruments to assess the current practices employed for teaching evaluation within Canadian university geography departments.

The purpose of this chapter is to provide an explanation for the research methodology employed in this project. The chapter commences with an introduction to the different data sources (Table 4.1). An explanation for the selection of the data sources follows. In Sections 4.3 and 4.4, the national level survey and oral interviews are described respectively. The description for each includes: participant selection, instrument design, survey administration and analysis employed.

Table 4.1: Data sources

Data Source	Number of Requests for Participation	Number of Actual Participants
Survey of geography departments	47	10
Interview with chairs of geography departments within Canada	40	23
Interviews with informants suggested by chairs of geography departments	21	11

Table 4.1 demonstrates that of the initial 47 geography departments that were surveyed only 10 returned surveys. This is a small response rate, but not problematic because it represented a broad group of geography departments. The respondents included: departments geographically spread out across Canada; departments that were classified as research-intensive and undergraduate-intensive, both small and large departments and ones located within Faculties of Arts and within Faculties of Science. Additionally, the written surveys were being used to inform the researcher about the type of questions to ask in the interviews. Seven fewer requests for participation were sent out for the oral interviews. This is because one department of geography (University of Windsor) closed during the period of this study and because of the researcher's inability to speak French any French speaking institutions were excluded. The response rate to the written survey was high (56%), but again does not represent a large group of individuals. This small sample size is justified because there is a finite population, the results were reproducible (i.e., a second survey with the same survey group would provide similar results), the content validity was high (i.e., the range of meaning of effective teaching attributes was

captured in the survey response), the construct validity was high (i.e., there was a logical relationship among the variables), the reliability was high (i.e., similar messages/themes were coming from the research participants) and the credibility of the participants was high because they are both practitioners of teaching and evaluators of teaching.

4.2 Site Selection and Sample Characteristics

In this section the reasoning used for the selection of geography departments within Canada for this study will be explained. In the second part of this section geography departments within Canada will be classified into three groups. Geography departments in Canada offer a wide variety of undergraduate degrees in Humanities, Social Sciences, Science and Environmental Science/Studies and at a variety of levels including bachelor's, master's and doctoral. A survey of the highest degrees offered by Canadian geography departments reveals that 24 departments offer the PhD degree as their highest degree, five offer the Master's degree as their highest degree, and 13 offer the Bachelor's degree as their highest degree (Table 4.2 and Table 4.3).

These departments do not all reside in the same Faculty within their institutions, and may or may not include the word 'geography' in the department name (Table 4.2). A review of the data in Table 4.2 reveals that 10 departments reside in Faculties of Arts, five departments reside in Faculties of Social Sciences, seven departments reside in combined Faculties of Arts and Science, five departments reside within Faculties of Science, four reside in Faculties of Arts and Social Sciences and eight reside within Faculties that are unique to their institution (e.g. Science and Environmental Studies). In Canada, 42 departments offer undergraduate geography degrees. 32 of these departments

are called 'Department of Geography', nine are called 'Geography and …' and one department does not include the word 'geography' in its name (Table 4.2).

Table 4.2: Summary of Canadian geography department including department name and the Faculty of residence

Institution[a]	Department Name[b]	Faculty[c]
University of Alberta	Dept. of Earth & Atmospheric Sciences	Science
Bishop's University	Dept. of Environmental Studies & Geography	Social Sciences
Brandon University	Dept. of Geography	Science
University of British Columbia	Dept. of Geography	Arts
Brock University	Dept. of Geography	Social Science
University of Calgary	Dept. of Geography	Social Science
Carleton University	Department of Geography & Environmental Studies	Arts & Social Sciences
University College of the Fraser Valley	Dept. of Geography	Arts & Applied Arts
Concordia	Dept. of Geography, Planning & Environment.	Arts & Science
University of Guelph	Dept. of Geography	College of Social & Applied Human Sciences
Lakehead University	Dept. of Geography	Science & Environmental Studies
Laurentian University	Dept. of Geography	Arts & Social Sciences
Université Laval	Dépt. de Géographie	De Foresterie et Géomatique
University of Lethbridge	Dept. of Geography	Arts & Science
Malaspina University-College	Dept. of Geography	
University of Manitoba	Dept. of Environment and Geography	Arts
McGill University	Department of Geography	Science
McMaster University	School of Geography & Earth Sciences	Science
Memorial University of Newfoundland	Dept. of Geography	Arts
Université de Moncton	Dépt. de Géographie	des Arts et Sciences Sociales
Université de Montréal	Dépt. de Géographie	des Arts et Sciences
Mount Allison University	Dept. of Geography & Environment	Social Sciences
Nippissing University	Dept. of Geography	Arts & Science
University of Northern British Columbia	Dept. of Geography	Arts
University of Ottawa	Dept. of Geography	Arts
Université du Québec à Montréal	Dépt. de Géographie	Sciences Humaines
Queen's University	Dept. of Geography	Arts & Science - Social Science
University of Regina	Dept. of Geography	Arts
Ryerson University	Dept. of Geographic Analysis	Arts

Institution[a]	Department Name[b]	Faculty[c]
Saint Mary's University	Dept. of Geography	Arts
University of Saskatchewan	Dept. of Geography	College of Arts & Sciences
Simon Fraser University	Dept. of Geography	Arts & Social Sciences
University College of Thompson Rivers (formerly: University College of the Cariboo)	Thompson Rivers Geography	Arts
University of Toronto	Dept. of Geography and Program in Planning	Arts & Science
Trent University	Dept. of Geography	
Trinity Western University	Dept. of Geography	Faculty of Humanities and Social Sciences
University of Victoria	Dept. of Geography	Social Sciences
University of Waterloo	Dept. of Geography	Environmental Studies
University of Western Ontario	Dept. of Geography	Social Science
Wilfrid Laurier University	Dept. of Geography & Environmental Studies	Arts
University of Winnipeg	Dept. of Geography	Science
York University	Dept. of Geography	Arts & Science and Engineering

Notes:

a. The definition being used for a department of geography in this research is any unit that offers an undergraduate degree in geography. Augustana College, Medicine Hat College and College of New Caledonia have been excluded from this list because although they offer university-level geography courses, they do not offer a degree.
b. The name of the department was determined from the Canadian Association of Geographers Annual Directory (Falcingo, 2006). The directory reflects information as of July 1, 2006.
c. The faculty in which the department resides was determined from each institution's website which was accessed on August 17, 2007. Some university websites do not provide information on the faculty in which the department resides. Departments for which the faculty in which it resides could not be determined from the website have been left blank.

It has been stated that the level of research-intensiveness at a university and within a department will impact undergraduate student learning as well as undergraduate teaching (Boyer Commission, 1998). The level of research-intensiveness within a department can be measured in a variety of ways including: identification of the highest degree offered by an institution (Table 4.3); the size of graduate to undergraduate enrolment (Table 4.3); the research productivity of the faculty members (e.g. number of publications/annum, research dollars/annum); or from published sources about the research-intensiveness of the institution (e.g., Maclean's Annual University Ranking). For the purpose of this research a department was considered research-intensive if it offered a PhD program and undergraduate-intensive if it did not offer a PhD program. This results in 24 departments being classed as research-intensive and 16 being classed as undergraduate-intensive.

Geography departments range in size across Canada from a department with only two faculty members to a department with 53 faculty members, from a department with 32 undergraduate program students (majors, honours, and minors) to a department with 1,024 undergraduate program students, and from a department with seven graduate students to a department with 152 geography graduate students (Table 4.3). The size of a geography department relative to the institution may also impact teaching (Table 4.3). These two factors can impact teaching by affecting class size, available resources, available teaching assistants, instructor teaching loads, and culture within the department. The number of courses offered is correlated to the number of faculty ($r=0.48$) in Canadian geography departments. There is wide variation in the number of courses/faculty member. For example, Brock University has the highest ratio with seven

undergraduate courses/faculty member; whereas, University of Toronto offers 1.8 courses/faculty member (Windsor had the lowest ratio of 1.2 courses/faculty member, but this number is not likely comparable given the flux that geography has undergone at Windsor in recent years. The number of courses per faculty member will affect the number of times an undergraduate student encounters a specific instructor which in turn may influence teaching and teaching evaluation practices.

Geography is a department that typically performs a lot of service teaching (i.e., instruction of students from programs outside of geography). This makes geography a good discipline in which to study because of the broad range of students in geography courses will likely make an evaluation scheme that is appropriate in geography to be appropriate in a wide variety of other disciplines.

Geography is a single discipline that provides the researcher with data applicable to a variety of post-secondary degrees, and faculties due to its placement within the institutional structures (e.g., Faculty of Arts vs Science). As well, it is a single discipline where the size of the department and level of research-intensiveness varies greatly from one institution to another. Thus, it is an excellent unit to assess measures of teaching effectiveness with the objective of having broad applicability to other disciplines in the future.

Table 4.3: A summary of the institutions that offer geography degrees, the number of faculty, the number of undergraduate geography students, the number of geography graduate students, the number of full time students enrolled at the institution and the number of undergraduate courses offered. (F/T = full time enrolment; u/g = undergraduate; grad = graduate; Inst = institution; B = bachelor's degree; M = master's degree; P = doctoral degree)

Institution[a]	# of Faculty [b]	U/G Geog Students[b]	Grad Geog Students [b]	F/T u/g Students [c]	F/T Grad Students [c]	Total F/T Enrolment (grad and u/g) [c]	# of u/g courses [d]	Student Ratio Geog Grad: Geog u/g	Student Ratio Inst grad: Inst u/g	Highest Level of Degree Granted[e]	Faculty Union-ization[f]
University of Alberta	38	83	107	28810	6882	35692	79	1.289	0.239	P	N
Bishop's University	7	105	0	2093	7	2100	41	0.000	0.003	B	Y
Brandon University	10	115	0	2275	4	2279	49	0.000	0.002	B	Y
University of British Columbia	28	382	93	26414	6882	33296	68	0.243	0.261	P	N
Brock University	12	444	0	13400	650	14050	84	0.000	0.049	B	Y
University of Calgary	22	325	152	20324	4323	24647	65	0.468	0.213	P	N
Carleton University	25	309	78	16900	2270	19170	60	0.252	0.134	P	Y
University College of the Fraser Valley	10	157	0	4719	2134	6853	47	0.000	0.452	B	Y
Concordia	13	399	63			0	83	0.158		M	Y
University of Guelph	18	376	43	17800	1900	19700	38	0.114	0.107	P	Y
Lakehead University	8	345	0	5600	450	6050	49	0.000	0.080	M	Y
Laurentian University	10	345	0	7200	300	7500	33	0.000	0.042	B	Y
Université Laval	16	304	0	19943	5629	25572	69	0.000	0.282	B	
University of Lethbridge	15	146	18	6860	183	7043	57	0.123	0.027	P	N
University of Manitoba	23	420	51	18772	2472	21244	95	0.121	0.132	P	Y

Institution[a]	# of Faculty [b]	U/G Geog Students[b]	Grad Geog Students [b]	F/T u/g Students [c]	F/T Grad Students [c]	Total F/T Enrolment (grad and u/g) [c]	# of u/g courses [d]	Student Ratio Geog Grad: Geog u/g	Student Ratio Inst grad: Inst u/g	Highest Level of Degree Granted[e]	Faculty Union-ization[f]
McGill University	25	195	60	19740	6075	25815	79	0.308	0.308	P	N
McMaster University	29	64	64	19400	2500	21900	82	1.000	0.129	P	N
Memorial University of Newfoundland	18	364	34	13325	1546	14871	53	0.093	0.116	P	Y
Université de Moncton	6			4680	426	5106	40		0.091		
Université de Montréal	20	271	90	20247	8112	28359	50	0.332	0.401	P	
Mount Allison University	7	102	0	2107	12	2119	39	0.000	0.006	B	Y
College of New Caledonia	2	409	0			0	8	0.000			Y
Nippissing University	11	348	0	3800	0	3800	58	0.000	0.000	B	Y
University of Northern British Columbia	11	108	12	2241	318	2559	55	0.111	0.142	P	N
University of Ottawa	19	314	34	24700	3000	27700	93	0.108	0.121	P	Y
Université du Québec à Montréal	25	992	80	18465	3643	22108	115	0.081	0.197	P	
Queen's University	17	362	58	14040	2900	16940	62	0.160	0.207	P	Y
University of Regina	13	114	11	8866	639	9505	51	0.096	0.072	P	Y
Ryerson University	19	186	22	15935	756	16691	71	0.118	0.047	M	Y
Saint Mary's University	8	98	0	6300	355	6655	54	0.000	0.056	B	Y
University of Saskatchewan	14	87	49	13401	1872	15273	67	0.563	0.140	P	Y
Simon Fraser University	25	493	52	11025	3046	14071	94	0.105	0.276	P	N

Institution[a]	# of Faculty [b]	U/G Geog Students[b]	Grad Geog Students [b]	F/T u/g Students [c]	F/T Grad Students [c]	Total F/T Enrolment (grad and u/g) [c]	# of u/g courses [d]	Student Ratio Geog Grad: Geog u/g	Student Ratio Inst grad: Inst u/g	Highest Level of Degree Granted[e]	Faculty Union-ization[f]
University College of Thompson Rivers	10	97	0	4719	2134	6853	37	0.000	0.452	B	Y
University of Toronto	53	925	144	53000	11000	64000	95	0.156	0.208	P	N
Trent University	10	379	13	6690	220	6910	39	0.034	0.033	B	Y
Trinity Western University	6	32	0	1894	317	2211	21	0.000	0.167	B	Y
University of Victoria	23	426	78	11127	2088	13215	79	0.183	0.188	P	N
University of Waterloo	22	418	66	18000	2500	20500	74	0.158	0.139	P	N
University of Western Ontario	26	313	45	26500	3700	30200	64	0.144	0.140	P	Y
Wilfrid Laurier University	20	283	43	11800	600	12400	99	0.152	0.051	P	Y
University of Windsor	9	60	7	12000	1220	13220	11	0.117	0.102		Y
University of Winnipeg	18	195	0	6462	0	6462	66	0.000	0.000	B	Y
York University	30	1024	51	39100	3300	42400	86	0.050	0.084	P	Y
Descriptive Information											
Minimum Value	2	32	7	1894	4	2100	8	0	0		
Maximum Value	53	1024	152	53000	11000	64000	99	1.289	0.452		
Range	51	992	145	51106	10996	61900	91	1.289	0.452		
Mean	17.5	307.5	39	14163	2350	16513	61	0.200	0.100		
Standard Deviation	9.7	230	41	10625	2538	12794	23	0.300	0.100		

Notes:

a. The definition being used for a department of geography in this research is any department that offers an undergraduate degree in geography. Augustana College, Medicine Hat College and College of New Caledonia have been excluded from this list because although they offer university level geography courses they do not offer a geography degree.
b. The number of faculty (including emeriti), number of geography undergraduate students and number of geography graduate students was determined from the Canadian Association of Geographers Annual Directory 2006 (Falcingo, 2006). The directory reflects the information as of July 1, 2006.
c. The full-time enrollment of undergraduate and graduate students was determined from the Association of Universities and Colleges (AUCC) website on November 26, 2007 (AUCC, 2007).
d. The number of undergraduate courses was determined from departmental websites on August 21 and 22, 2007 and reflects the 2007-08 undergraduate calendar.
e. The highest degree offered is P – PhD, M- Masters, and B- Bachelors. The highest level of degree offered was determined from the Canadian Association of Geographers Annual Directory and reflects information as of July 1, 2006 (Falcingo, 2006). Moncton, New Caledonia and Windsor did not provide a level of degree in the CAG Annual Directory so have not been included for this work.
f. Union status was obtained from Collective Agreements accessed from University web sites on June 13, 2007 and confirmed with Phyllis DeRosa Koetting the Executive Director of the McMaster University Faculty Association in November 2007.

4.3 National Level Survey of Geography Departments

a. Introduction

The national-level survey was designed to provide the researcher with a preliminary understanding of the breadth of teaching evaluation practices currently being used in Canadian geography departments. The survey provided additional information about the demographics of the departments. The knowledge gained from the national-level survey was used to inform the methodology for the second part of this thesis (see Sections 4.4 and 4.5).

b. Participant selection

All university geography departments within Canada were invited to participate. Departments were identified from the Canadian Association of Geographers Annual Directory (Falcigno, 2005)

c. Instrument design

The survey instrument received ethics approval from Wilfrid Laurier Research Ethics Board in August 2005 (Appendix One). The survey consisted of 32 questions and required approximately 60 minutes to complete (Appendix Two). The survey instrument was divided into six sections, as follows:

- student enrolment information including both graduate and undergraduate information;
- undergraduate course information including the number of courses offered, the rank of the individual teaching the courses and the availability of on-line courses;
- instructional staff information including rank, gender, unionization about both teaching and teaching assistant staff;
- teaching evaluation information including reasons to evaluate, tools used to evaluate and evaluation process; and,

- chair or designate perspectives on the teaching evaluation processes.

At the conclusion of the survey instrument a request was made for submission of additional documents about teaching including: course evaluation instrument, Collective Agreements, faculty handbooks, and senate and departmental policies on teaching evaluation.

d. Survey administration

A total of 47 surveys was mailed via Canada Post to each chair of a geography department within Canada that offered an undergraduate degree as of September 2005. The surveys were followed up by two subsequent email reminders. The second email reminder also included a digital copy of the survey instrument to facilitate electronic completion of the survey or alternatively printing the survey and completing if the original survey had been misplaced. Survey responses could be submitted by return post (a postage paid envelope was provided) or electronically via email.

e. Analysis

Quantitative data analysis was completed using SPSS. The intent was to use NVivo (a qualitative research and data analysis software package) to assist with the qualitative data analysis. The small survey response (10 surveys) coupled with minimal comments in response to the open-ended questions resulted in NVivo being an unnecessary tool to review the small sample.

4.4 Oral Interviews

a. Participant Selection

Two sets of interview participants were selected in this work. The first set of interview participants was selected on the basis of their role as chair of a university geography department within Canada. Due to an inability of the researcher to communicate in French, only chairs of English-speaking universities were contacted.

The second set of interview participants were identified by the chairs based on their response to the question, "Is there anyone else at your institution that it would be helpful for me to contact to more fully understand the teaching evaluation process?" The second set of interview participants consisted of faculty members within the chairs' geography department, senior administrators (e.g., Faculty deans), faculty members from other academic units, a student leader, and staff from teaching and learning centres.

b. Instrument design

The survey instrument received ethics approval from Wilfrid Laurier Research Ethics Board in June 2007 (Appendix Three). The survey consisted of 10 questions and required approximately 30 minutes to complete (Appendix Four).

c. Survey administration

A total of 40 chairs was contacted via email in June, 2007 with a request to participate in an interview. The initial email also contained a copy of the ethics consent. A research assistant followed up by telephone to set up a convenient interview time. The research assistant made a minimum of three calls to each chair in an attempt to set up interview appointments. Twenty-three chairs (58%) agreed to participate and interviews were conducted between June and November, 2007. The reason provided by the 17

individuals who elected not to participate in the study included: being too busy (1), the department of geography no longer existed at the institution (1) and several never returned the researcher's email or research assistant's phone calls (15).

The final question in the interview with the first 10 chairs was to identify additional participants to interview. The additional participants included: faculty members within the chairs' geography department; senior administrators (e.g., Faculty deans); faculty members from other academic units; student administrative bodies; and, staff from instructional development centres. Although the second set of interview participants were not specifically asked for additional participants some provided participants. These individuals were contacted and invited to participate. The invitation to participate included an email from the researcher which included the ethics consent and a follow-up phone call (minimum three) to set up a convenient interview time.

A total of 21 individuals was contacted between July and October, 2007 and invited to participate in the interviews. Eleven agreed to participate and interviews were conducted between July and November, 2007. The reason provided by the 10 individuals who elected not to participate in the study included: being too busy (3); not interested in participating (2); and, several never returned the researcher's email or research assistant's phone calls (5).

d. Analysis

All of the surveys were tape recorded, transcribed and entered into NVIVO, a qualitative data management computer program. The data were coded by the researcher. In an attempt to ensure coder reliability, three of the interviews were provided to another researcher to code. Codes were assessed for consistency between the two researchers.

Coding was consistent in 97% of the codes. This is a high, but not surprising, coding consistency given the strong correlation between the research objectives, interview questions and code tree. A total of 149 codes was created in NVIVO. These codes align into seven themes: interviewee characteristics, evaluation method, purpose of evaluation, excellence rewards, quality improvement, evaluation process effectiveness, and good teaching definition (Appendix Five). Thematic analysis was completed of the data. The themes were tightly aligned with the research objectives which were in turn tightly aligned to the node tree.

CHAPTER FIVE:

Defining Good Teaching

5.1 Introduction

The results and analysis from the interviews of geography chairs, student leaders, teaching and learning centre staff, geography instructors and university administrators to address the first and second research questions:

> ***What is effective teaching?*** *and,*
> ***What is effective teaching within the discipline of geography in higher education?***

are presented in this chapter. The literature that was reviewed in Chapter Two informed the creation of a conceptual framework to describe effective teaching in general and specifically within the discipline of geography (Figure 2.1). The fit between the literature informed conceptual framework and the data collected in this research will form the analysis component of the following chapter.

5.2 Effective Teaching

When chairs of geography departments were asked, approximately at the mid-point of a thirty-minute telephone interview, how they would define good teaching, they all paused for reflection (see Appendix Four for the interview questions). This was the

only question in the interview that consistently resulted in the pause. This pause may reflect the complexity of the response that would be required to answer the question and the nearly non-existent amount of time that exists in a telephone interview for a detailed and thoughtful response to the question. Alternatively, it may reflect the interview participants' desire to give the 'right answer' or 'acceptable answer' about a question that they may only occasionally consider. Nearly one-half of the chair participants began their response by saying that it was a good question or that it was a tough question. A typical response began,

> "That's a very good question. It isn't one of those things that you can put your finger on?"

Several participants requested clarification if the researcher wanted the participant's opinion or the institution's opinion. In nearly all instances the chairs provided detailed characteristics of good teaching:

> "I think that number one, you've got to have good classroom prep. That means different things to different faculty members. .... I spend a lot of time trying to keep my lectures current. I try to be topical so that ... I make sure that I try to bring that into the lecture. I try to make the material relevant. That's an important thing that we need to do is to show what we're doing is relevant in some way or another. I also think that, along with the enthusiasm that you bring to the class, you also have to maintain consistency within teaching. That means you don't cut a lot of side deals with students.... Inconsistent marking, inconsistent examinations, whatever, are really bad in teaching. Feedback to students – I find that for some faculty feedback to the students is appalling. When you have a written paper, for example, not to provide detailed written comment on it at the end of the paper so the student knows where they went wrong or where they went right is wrong. We have to do that - feedback and constructive feedback. Not that "you are the biggest dolt in the world and you will never pass this course". It needs to be constructive. There are so many obvious things – what makes a good teacher. There is also something intangible about a good teacher. Some people are good teachers. They just have a way of being organized, presenting things in a clear and concise manner, in a friendly, non threatening manner."

The 23 chairs that responded identified a total of 15 characteristics and 3 meta-characteristics of good teaching in their definitions (Table 5.1). Only seven chairs (i.e., 29%) explicitly associated good teaching being reflected in enhanced student learning, although 11 of the 24 chairs (i.e., 46%) identified engagement as a characteristic of good teaching. A number of the other characteristics that were identified by chairs would logically lead to student learning and engagement.

Table 5.1: A summary of the results from asking the chairs of Canadian geography departments to define good teaching compared with the best practices in undergraduate teaching proposed by Chickering and Gamson (1987) and the attributes of effective teaching proposed in Figure 1.1. The shaded areas represent meta-characteristics of effective teaching.

Good Teaching Characteristics	Overall		Research		Under-graduate		Unionized		Non-unionized		Chickering & Gamson Best Practices	Conceptual Framework Parameters of Effective Teaching (Fig. 2.1)
	# of Chairs	%	# of Chairs	%	# of Chairs	%	# of Chairs	%	# of Chairs	%		
Student engagement	15	68	8	53	7	100	9	64	6	68		
Knowledge transfer	6	27	4	27	2	29	3	21	3	33		
Student learning	7	32	5	33	2	29	2	14	5	56		
Enthusiasm	4	18	2	13	2	29	4	29	0	0		Delivery of Instruction
Skill development	4	18	1	7	3	43	3	21	1	11		Skill Development
Challenging	4	18	4	27	0	0	0	0	4	44	Communicates high expectations	Discipline Knowledge
Course organization	3	14	3	20	0	0	3	21	0	0		Course Organization & Administration
Consistency	2	9	1	7	1	14	2	14	0	0		Student/Instructor Interaction & Assessment Tasks
Currency of material	4	18	4	27	0	0	2	14	2	22		Discipline Knowledge
Student assessment	2	9	2	13	0	0	1	7	1	11	Prompt feedback	Assessment Tasks
Knowledgeable	5	23	4	27	1	14	2	14	3	33		Delivery of Instruction
Relevancy	2	9	1	7	1	14	2	14	0	0		Discipline Knowledge
Student & instructor interaction	2	9	1	7	1	14	2	14	0	0	Student faculty contact	Student/Instructor Interaction & Assessment Tasks
Audience awareness	1	5	0	0	1	14	1	7	0	0	Diverse talents and ways of learning	Delivery of Instruction

Good Teaching Characteristics	Overall		Research		Under-graduate		Unionized		Non-unionized		Chickering & Gamson Best Practices	Conceptual Framework Parameters of Effective Teaching (Fig. 2.1)
	# of Chairs	%	# of Chairs	%	# of Chairs	%	# of Chairs	%	# of Chairs	%		
Feedback	1	5	1	7	0	0	1	7	0	0	Prompt feedback	Assessment Tasks
Communication	3	14	2	13	1	14	1	7	2	22		Delivery of Instruction & Student Instructor Interaction & Assessment Tasks
Classroom experience	2	9	2	13	0	0	1	7	1	11		Delivery of Instruction
Pedagogies	2	9	2	13	0	0	0	0	2	22		Professional Development

Interestingly, only four of the seven good practices identified by Chickering and Gamsen (1999), were described by geography chairs when asked about their definition of good teaching. The three that were not mentioned include: good practice encourages

- cooperation among students
- active learning; and,
- emphasizes time on task.

Two of the above practices (i.e., active learning and cooperation among students) are described by chairs when asked about good teaching within geography (see Section 5.3). Engagement of students is thought to enhance student learning in part by an increased amount of time that is spent on the subject materials (Carini et al., 2006).

When chair responses are compared to the conceptual framework developed by the researcher, based on the literature (Figure 2.1), all of the parameters were identified by the chairs in their definitions of good teaching. Professional development, although not mentioned in the discussion with chairs about good teaching or good geography teaching, is described by the chairs when identifing methods to reward teaching excellence and enhance teaching quality are described later in the interview.

Table 5.1 illustrates a tight correlation between effective teaching attributes identified by research-intensive departments and those identified by non-unionized departments. None of the undergraduate-intensive departments that participated in this research had non-unionized faculty. The two most commonly identified attributes of effective teaching in undergraduate-intensive departments and departments with unionized faculty were enthusiasm and skill development. These two attributes scored quite low in research-intensive and non-unionized departments. The three attributes most

commonly identified in research and non-unionized departments were currency of material, knowledgeable instructors and challenging material.

Four university administrators were interviewed at the suggestion of geography chairs. All four administrators provided definitions of good teaching that were tightly aligned with the concept of student engagement.

> "To me, good teaching is the ability to engender in the learner the wish to want to know more. Sometimes that's because they learn facts that turn them on and make them really excited. Sometimes it's because students fall in love with your passion for your subject."

One of the administrators described the link between student engagement and the characteristics of good teaching that are described in Table 5.1 and Figure 1.1.

> "For me, good teaching has to have a very high level of engagement on the part of the instructor and the student. There's a sense of process that involves shared learning. That may involve developing mutual understanding of a course and its objectives, having those articulated clearly, having students involved in the process perhaps of even defining the objectives for the course – so it's not just a unidirectional thing coming from someone supposedly charged with managing a curriculum – but finding means for student engagement in that process. That can boil right down to the curriculum design in terms of types of assignment, student involvement in the classroom as teachers – I find we do a lot of that. We really work very hard to ensure that our own students become active in the teaching process."

This same administrator introduced the idea of scholarly teaching and suggested that good teaching requires an awareness of the scholarship of teaching and learning.

> "The second element would be an awareness of the scholarship of teaching and learning. That comes from that reflective process and it leads people to think about where their teaching methods may be dying. It is a trajectory and it is a process that people are engaged in. Those are probably the two main things if I was to keep it brief – engagement and understanding a connection to the scholarship of teaching and learning."

It is interesting to note that administrators (i.e., deans, associate deans and vice-presidents) did see a component of professional development (i.e., connection to the scholarship of teaching and learning) as part of the definition of good teaching. The breadth of response that is encountered when asking participants to define good teaching supports the observation that teaching is a complex multi-faceted construct (Marsh and Roche, 1997; Cashin, 1989; Marsh, 1987). All chairs either specifically identified that effective teaching enhanced student learning (32%) or identified characteristics of teaching that have been demonstrated to enhance student learning in their definition. This supports Ramsden (1992) who defined good teaching as that which results in the greatest student learning. Several chairs addressed the idea that individual instructors have different abilities, skills and preferences. As suggested by Braskamp et al. (1984), instructors should be encouraged to teach to their strengths. One chair summarized this concept particularly well.

> "Some faculty are more comfortable with the smaller classrooms and do excellent at that but they're not comfortable with the large first-year classes. ...you might try and put faculty in environments where they're optimized."

Administrators appear to have a stronger sense of the idea of engagement being linked to student learning than chairs. This may be a function of having had the opportunity of more experience, a broader understanding of different disciplines and their approaches to good teaching or increased exposure to tools like NSSE (National Survey of Student Engagement) which increasingly Canadian universities are participating in annually. This stronger sense of understanding the ideas of engagement may also derive from being

more reflective about their own teaching or from having to more commonly answer questions about teaching practices.

The breadth of ways in which good teaching is defined may have a major impact on teaching evaluation and best practices. How can teaching be effectively evaluated if it is unclear what is good teaching? It seems reasonable to assume if 24 chairs of geography have a very broad and sometimes not consistent definition of good teaching that university students at different positions in their degree and at different institutions would have very different definitions.

5.3 Effective Geography Teaching

A subset of geography chairs (13 of 23) were asked if there was anything that was geography-specific about good teaching (i.e., was there anything that needed to be added to the definition of good teaching to define good teaching in geography). All the chairs spoke of some aspect of teaching that was geography-specific. This supports the idea that teaching consists of a suite of effective teaching skills and techniques that are discipline-specific (Crebbin, 1997; Ramsden, 1991; Murray and Renaud, 1995; Shulman, 1993; Sullivan and Skanes, 1974). The chairs identified four additional characteristics of effective teaching as being geography-specific:

- opportunities for experiential learning (i.e., field and lab);
- the ability to visually present and interpret data with students.
- the ability to help students understand the global and interdisciplinary nature of the subject material; and,
- the need for technological currency.

Not surprisingly, the geography chairs identified four of the same characteristics of effective geography teaching that are described in the literature.

- preparing for and conducting numerous and extended field trips (Tricart, 1969; Gold et al., 1991; Abler, 1994; Cooke, 1998; Chalkley et al, 2000);
- teaching topics that require computer-assisted teaching and learning e.g., GIS (Gold et al, 1991; Chalkley et al., 2000); and,
- the interdisciplinary nature of geography (i.e., blending science, social science and humanities as well as 'borrowing' methodologies from a wide variety of other disciplines and linking theoretical and applied aspects), (Tricart, 1969; Abler, 1994; Marantz and Warren, 1998; Farrington, 2000; Geography Benchmarking Group, 2000).

The one item that was not identified by chairs of geography as a characteristic of good teaching in geography, but has been identified in the literature is:

- a large component of teaching that involves instruction of audiences beyond the traditional tuition-paying students due to the strong sense of community responsibility and outreach (i.e., a large civic responsibility) (Abler, 1994).

The questions that were asked in the survey prior to the good teaching question may have guided the chairs to think only of teaching within the context of university students. Alternatively, thirteen years have passed since Abler identified the idea of a civic responsibility. It may be that the increasing financial constraints on universities, and increasing demands on instructors' time in recent years, has resulted in a focus on internal teaching more than on external teaching.

In Chapter Six the results and analysis of teaching evaluation practices in Canadian university geography departments will be described.

CHAPTER SIX:

Teaching Evaluation Practices

6.1 Introduction

In Chapter Three the methods used to evaluate teaching in higher education were explored to inform an answer to the third research question:

How is teaching evaluated in higher education?

and to develop a conceptual framework of teaching evaluation practices. In the current chapter responses from research participants will be analyzed to explore the fourth research question:

What is the breadth of teaching evaluation practices currently used in Canadian university geography departments?

The responses from research participants regarding current teaching evaluation practices will be compared to those described in the review of general teaching evaluation strategies provided in Chapter Three and will be used to test the conceptual framework provided in Chapter Three. Research participant responses to how results of teaching evaluations are used to reward and enhance teaching quality will be explored in the current chapter. As well, research participants will provide input to understanding the

effectiveness of the current teaching practices. These data will be used to explore research questions five, six and seven provided below:

> ***How are the results of teaching evaluation used to enhance teaching quality within Canadian geography departments?;***
>
> ***How are the results of teaching evaluations used to reward teaching excellence within Canadian geography departments?*** **; and,**
>
> ***How effective are current teaching evaluation practices?***

The responses to these questions will inform the current chapter and provide the data necessary for the creation of a revised conceptual framework of teaching evaluation practices and the creation of a teaching evaluation system model.

6.2 Teaching Evaluation Practices: Canadian Geography Departments

Prior to reviewing the evaluation practices within Canadian geography departments, it is important to understand the reasons why teaching is evaluated in these departments. Chairs identified four reasons to evaluate teaching. The first reason surrounded issues associated with departmental and institutional accountability. Over 80% of chairs identified that teaching was evaluated in order to demonstrate evidence that teaching was regularly measured as an accountability indicator. In other words, teaching was being evaluated for a summative purpose.

The second reason that chairs identified for evaluating teaching was in order to provide instructors feedback on their teaching (55% of chairs). This was seen as formative purpose for evaluating so that instructors had the opportunity to reflect on their strengths and weaknesses with the intended outcome that this reflection would lead to

future enhancements in their teaching. Providing formative feedback on teaching quality has long been seen as a purpose of teaching evaluation (Murray, 1980; Marsh, 1987; Marsh and Roche, 1997).

The third reason also had a summative, administrative purpose. Teaching was evaluated in order to provide evidence for administrative decisions regarding tenure, promotion and, at some institutions, merit increases (50%). This purpose has been identified in the literature as a measure to be used for administrative purposes to assist in guiding their decisions about promotion, tenure and salary (Murray, 1980; Marsh, 1987; Marsh and Roche, 1997).

The fourth reason provided by chairs for evaluating teaching was to gain the student voice and in some cases this information was used by prospective students in order to assist them in making decisions about future course enrolment (21%). This reason was met with some concern by chairs because it was perceived that, because of the public sharing of the results, some of the most useful data on effective teaching may not have been captured (i.e., students may not have been as forthcoming because of the impact of their evaluations becoming public, even though they would remain anonymous). The chair who suggested this felt that part of the reason that students may perceive this concern is that they are a small institution, will be taught multiple times by the same instructor and see the instructor in very human terms as a person. Murray (1980) and Marsh (1987) identified gathering information to assist prospective students in their course selection as a purpose for evaluating teaching.

The literature identified two additional purposes for evaluating teaching effectiveness that were not identified by any of the chairs of geography. Those two purposes are:

- to assess the quality of individual courses to be used for course and curriculum improvement and design (Marsh, 1987); and,
- to provide data for research on teaching (Murray, 1980; Marsh, 1987).

It is unclear why these purposes were not identified by chairs. Perhaps it is because course and curriculum improvement and design is often completed by a committee of geographers who would be participants and colleagues of those being evaluated and that evaluation data is considered confidential between the administration and the instructor in many instances. As well, it may be that curriculum design follows more informal methods and is not based on content from course evaluations. It may not have been identified as being useful to these purposes because some chairs had concerns about the reliability and validity of the data. This was particularly the case at institutions/departments where on-line course evaluations had been introduced and the response rates had dropped considerably. At one institution response rates had dropped to less than 10%.

None of the chairs and only two of the administrators mentioned pedagogical research and none of the research participants described a purpose of teaching evaluation associated with providing data for research on teaching.

There are three key times in an instructor's career when teaching is evaluated within nearly all Canadian geography departments. The first time an instructor has their teaching evaluated is at the time of hiring. The second and most frequent occurrence of

evaluation occurs at the end of each course. The third occurrence is when an individual is due for a tenure and/or promotion review. The first and third occurrences are holistic evaluations whereas the course evaluation is a snapshot of one course at one point in time. The practices at each of these three occurrences will be described below.

a. Time of hiring

At the time of hiring a variety of approaches is used to assess teaching including: previous experience that may come from prior teaching (>90%) or through participation in courses/workshops/presentations about teaching and learning, teaching statement or philosophy (36%), sample presentation (64%), reference letters (100%), round-table dialogue (10%), and undergraduate student feedback (18%). If the candidate has any previous teaching experiences ,course evaluation data will be reviewed. As well the candidate's curriculum vitae (CV) will be reviewed for evidence of a commitment towards undergraduate teaching in the form of participation in workshops/courses/presentations related to teaching and learning. In some cases, this information is reviewed only from the CV; alternatively it is reviewed as part of a teaching dossier. A second approach that is used to evaluate teaching effectiveness at time of hire is a written teaching statement or teaching philosophy. The third approach is evaluation during an actual presentation. This may involve having the candidate present to an actual class, present to a 'mock class', or having teaching potential evaluated as part of the candidates research talk. Several chairs commented on the challenges of having a candidate present to an actual class. A fourth approach to evaluating competence is the evaluation of reference letters. These letters are typically reviewed for evidence of teaching abilities. A hypothesis when this research project was initiated was that there

would be a different approach taken by research-intensive schools and undergraduate-intensive schools when candidates were interviewed for tenure-track positions. The hypothesis was that teaching would matter more in the hiring process at primarily undergraduate schools and that research would matter more at research-intensive schools. None of the undergraduate-intensive departments described a process for obtaining the student voice on the potential to hire a candidate whereas 27% of the chairs of research-intensive institutions described a process for capturing the undergraduate student appraisal of potential candidates. The following quotes from chairs of research-intensive schools would certainly indicate that a high value is placed on teaching at the time of hire.

> "Teaching ability is one of the main criteria for hiring."
>
> "…teaching is an important part of a university professor's attributes."

A chair at a primarily undergraduate institution was much more pragmatic in the type of candidate that they would hire.

> "At hiring, you need some evidence that someone going into a primarily teaching position can teach. Beyond that, we're looking for some evidence of improvement. You can't expect someone to be great from the beginning, necessarily."

Two institutions, both research-intensive, have a form of a round-table discussion with potential candidates. In both cases the round-table is attended by faculty and the members of the selection committee with the stated of objective to engage the candidate in dialogue in order to ascertain:

> "…[an] impression of what their attitude towards teaching and students is."

> "There we'll ask about teaching, and how would you teach large classes, and what would you teach, and so on. Those things, I think, give one a fair indication of the teaching ability of the individual."

Only two institutions, both research-intensive, described a formalized procedure in which a potential candidate would meet with a group of undergraduate students as part of the interview process and the students would formally present feedback to the selection committee about their reflections on the suitability of the candidate.

The objective at the time of hiring consistently appears to be to hire a candidate that will initially be competent in the classroom and has the potential for future improvement in their teaching. This potential appears to primarily be measured in the form of attitude towards teaching, learning and students.

b. Course evaluations

The second time when teaching is normally evaluated is through course evaluations. In nearly all geography departments within Canada (>85%), course evaluations are completed at the end of each course. Often there are minimum class size requirements in order to ensure student anonymity when completing the evaluations. At the three institutions that do not complete evaluations following every course, they are completed on a regular cycle; typically once every three times the instructor teaches the course. The requirements for these institutions are described in their respective Collective Agreements. In all cases these geography departments reside in universities where teaching staff are unionized. Table 6.1 summarizes the process for course evaluations within the 23 geography departments within Canada that participated in this research.

Table 6.1 Summary of the course evaluation processes within Canadian geography departments

		Overall		Research		Undergraduate		Unionized		Non-unionized	
		# of Chairs	%	# of Chairs	%	# of Chairs	%	# of Chairs	%	# of Chairs	%
Instrument set by	Institution	15	68	8	53	7	100	11	79	4	44
	Faculty	6	27	6	40	0	0	1	7	5	56
	Department	1	5	1	7	0	0	1	7	0	0
Instrument delivered by	Students	7	32	6	40	1	14	4	29	3	33
	Teaching Assistant or Graduate Student	7	32	4	27	3	43	3	21	4	44
	Department support staff	1	5	0	0	1	7	1	7	0	0
	T&L centre staff	1	5	0	0	1	14	1	7	0	0
	Other institutional support staff	1	5	1	7	0	0	1	7	0	0
Format	On-line	4	18	3	20	1	14	2	14	2	22
	In-class	18	82	12	80	6	86	11	79	7	78
Access to results	Instructor	22	100	15	100	7	100	13	100	9	100
	Chair	16	73	12	80	4	57	10	71	6	67
	Dean	7	32	4	27	3	43	6	43	1	11

The majority of teaching evaluation instruments used in Canadian geography departments are used on an institution level (68%). A number of commercially available student rating forms exist and were described as being used at three institutions within Canada (e.g., CIEQ (Course Instructor Evaluation Questionnaire) form, IDEA (Individual Development and Educational Assessement) form, SEEQ (Student Evaluation of

Education Quality) form). The opportunity for variability between instruments is quite large. A study by Arreola (2000) has identified 520 rating items that could be used on course evaluations.

There is general consistency across the country that the instructor should be at arms length from the collection of course evaluations and should not have access to the results until after the end of term. Most often, other instructional staff from the geography department (e.g., graduate students, teaching assistants, and other faculty) administer the collection of data. Interestingly, at only one institution are the teaching and learning centre staff involved in the collection of data. It appears that teaching and learning centres perceive their role as being supportive rather than contributing to the administrative/accountable purpose that is often perceived from course evaluations.

Only four institutions reported collecting the data using an on-line form. Two of the four schools reported a significant drop in the response rate when they went to an on-line form. The third school was still in the beta testing stages and had not yet collected any data, although preliminary testing had indicated that they did not anticipate a decline in the response rate. Interestingly, at one of the schools where on-line forms had been introduced, the chair had collected some statistics on the response rates. The data showed a "really big drop" in response rate initially that has been followed with a slight increase. Even in classes that experienced a 50%+ decline in response rates, the t-tests on the mean score and the standard deviation were not significantly different. A second school that has gone to on-line evaluations reported response rates in the single digits (i.e., less than 10%) and were very concerned about the implications that the decline meant for capturing the student voice and for providing formative feedback to the instructor.

All of the course evaluation instruments collect formative feedback. At some institutions, summative data are also collected. The formative feedback occurs most frequently with the open-ended comments which students had the opportunity to add to more than 75% of the evaluations. Often these comments are available only to the instructor and are not available to administration (either the chair or the dean). In this way the feedback can only be used by the instructor to enhance subsequent offerings of the course. At some institutions, additional questions can be added to the instrument. These additional questions are designed by the instructor to solicit formative feedback to get a more complete picture of the student experience and to provide data on how to enhance the course in the future. A few institutions have standard multiple choice questions that ask for data that can be used to enhance the course in the future. At these institutions, where these data are not available to administration, it would be considered formative in nature.

In all cases the instructor of the course has access to the results following the end of the academic term. In some institutions this access does not occur until the end of the academic year. At most institutions the chair (73%) and at some institutions the dean (32%) has access. Surprisingly, there is not a lot of variation between unionized and non-unionized work environments, other than in non-unionized environments the dean is not often described as having access to the results from course evaluations on a regular basis (i.e., at the end of every course).

At all institutions, the results from course evaluations are used in the tenure and promotion process. At some of the institutions, the results are used in the annual merit increase. This is where the greatest distinction between unionized and non-unionized

instructional staff occurs. In the unionized environment, very few institutions have a merit increase and the merit increase is not set to reward excellence but is used more as a punitive measure, although it is rarely used that way either.

> "...we have performance increments. The union negotiates annually, I guess, or for several years in a row a cost of living increase. Pretty much every faculty member receives that and then a satisfactory performance increment. Generally I would say, in my time [as chair], I've not had to withhold performance increment."

How the results are used to enhance teaching quality and reward teaching excellence will be discussed in Section 6.3.

c. Merit, tenure and/or promotion

The third time in an instructor's career when teaching is often evaluated is for the awarding of tenure and promotion. For the awarding of tenure and promotion, all chairs assured the researcher that teaching was an important component of the decision. Generally, a combination of some form of teaching dossier or portfolio is prepared for use by the committee reviewing the candidate for tenure and/or for promotion. The teaching dossier would contain summative information from course evaluations. At some institutions there is also a colleague evaluation component. In some cases, this involves a chair observing a class or an identified excellent teacher within the candidates department evaluating their teaching. One of the geography professors interviewed for this research, but not a chair of a geography department, highlights some of the challenges they have encountered in evaluating their peers.

> "There's peer evaluation that goes on for promotion and tenure and renewal of probationary term contract. I've done a lot of them. I don't think they work particularly well. The reason I say that is – I think I give good feedback, it's not that I don't give good feedback – but it's unreasonable to ask a colleague, who's not anonymous and who's potentially even to some extent a friend or a mentor or something, it's completely unreasonable to expect them to give a poor teaching evaluation to a colleague. The double-blind peer review has worked for research; it simply is not replicated here in teaching. I have written letters for all sorts of folks. The details I give, I think, are honest, useful to the individual, and useful to the institution as far as outlining the strengths and weaknesses. But I can assure you the overall tone of each of those letters that I've written has been positive. The reason for that is that you're not about to end someone's career because you popped into a classroom and you saw it not going well. You may suspect that there's a problem. You may have heard there's a problem. But before I would write a letter, I would want to be in that person's class once a week for a year. And then, maybe, I would come out and say, 'Ok, I'm prepared to say this person's teaching is below standard.' There's no way anybody will do that on the kinds of peer evaluations that are being done. And to make it more onerous for the people who are doing the peer evaluations doesn't seem like the solution, either."

From the perspective of this research participant it appears that the culture within a department or within a university is critical to the value placed on teaching. This research participant also identified that there is variation in the culture among departments which in turn will impact the value placed on teaching.

> "But, that said, different departments have different cultures. In some departments teaching has a very small standard deviation so the people who perform excellently in the classroom and do a lot of teaching may end up getting a 1.75 out of 2. And the people who do a virtually terrible job all the time might get a 1.25 by virtue of number alone. That's a very small variation. Whereas, in other departments they are happy to give out 0.5 out of 2 for teaching or 2 out of 2 for teaching."

One of the initial hypotheses when this work commenced was that there would be a clear distinction in the culture around teaching between research-intensive and primarily undergraduate institutions. The interviews with geography chairs did not support this hypothesis. The intra-group variation is at least as large as the inter-group variation. The following two quotes, both from chairs of geography departments, highlight the

differences in culture at the two institutions. The first is from a primarily undergraduate department and is in response to the interviewer asking about merit increases.

> "In the department itself we look at teaching, research, and service. I think – this is going to be no surprise – the research file is the one that's going to take precedence."

The second quote comes from a chair at a research-intensive school and is in response to the interviewer asking about institutional approaches to ensure quality in undergraduate teaching.

> "It tries to convey quite clearly that it sees teaching as equally important as research in terms of faculty member activity. I'm not sure that that message is getting through necessarily to a lot of people, but I don't think the institution can be faulted for that. The message is quiet clear if you listen to it. … I think the important thing is to really create an environment where people become aware from the very first time they have contact with it that teaching is seen as important and a valued activity not just as something that you have to do in addition to your research. If you create that climate then, generally speaking, people will take it seriously. And if they run into problems then they'll realize very quickly that they're anomalies and they need to not be anomalies if they want to progress the way they'd like."

6.3 Rewarding and Enhancing Teaching

a. Rewarding excellence

After describing the perceived purpose of teaching evaluations and the mechanisms used to evaluate teaching, the participants were asked to describe what was done with the results of teaching evaluations to reward and enhance teaching quality. Table 6.2 summarizes the rewards for teaching excellence that were identified by chairs of geography departments. All chairs described teaching awards that were used to reward excellence in teaching. These awards were offered at a variety of different levels including departmental, faculty, institutional, provincially and nationally. As well, some

chairs described teaching awards that were presented by the undergraduate student association. There was a range of understanding about the awards that were available within the chair's institution.

> "Annually there is the medal for teaching excellence given at convocation. Somehow someone gets nominated for that and someone is chosen. I would assume teaching evaluations might play a role in that. I don't think it's necessarily the only thing."

> "We have teaching awards, many teaching awards. We have faculty teaching awards, we have university teaching awards, and external teaching rewards that we would nominate our outstanding instructors for. I'd say, at least, every year to two years we have somebody from the department getting one of those kinds of rewards."

Interestingly, more than 90% of the chairs at research-intensive universities, but only 29% of chairs at undergraduate-intensive universities, described merit increases as a reward for teaching excellence. Several of the undergraduate-intensive universities hired their teaching staff on contract basis and did not have a merit system in place. Four of the research-intensive universities described teaching chairs/fellowships/teaching release time as a reward for excellence in undergraduate teaching. This follows closely the model of research chairs and fellowships. It may be linked at these institutions as an attempt to 'value' undergraduate teaching and demonstrate a commitment to a culture of undergraduate teaching excellence. Three chairs identified teaching as an activity that was intrinsically rewarding when done well and saw that as a reward for teaching excellence. One hundred percent of chairs in departments with non-unionized faculty described merit increases, but only 50% of chairs in departments with unionized faculty described merit increases.

Finally, three chairs, all at research-intensive universities, identified tenure as a reward for teaching excellence. None of the chairs interviewed suggested that tenure would be successfully obtained based strictly on outstanding teaching. Although several pointed out, however, that they had a teaching stream of appointment that individuals who wished to focus on teaching only could choose to pursue.

There was also recognition on the part of three chairs that rewards had to be both intrinsic and extrinsic in nature. The use of awards would be an example of an extrinsic reward. Intrinsic rewards are more personal in nature, but for some individuals the opportunity to mentor a junior instructor might provide an intrinsic reward.

Table 6.2: Rewards for teaching excellence identified by chairs of Canadian geography departments

Reward	Overall		Research		Undergraduate		Unionized		Non-unionized	
	# of Chairs	%	# of Chairs	%	# of Chairs	%	# of Chairs	%	# of Chairs	%
Awards	22	100	15	100	7	100	13	100	9	100
Merit Increase	16	73	14	93	2	29	7	50	9	100
Teaching Chair, Fellowship, Teaching Release	4	17	4	27	0	0	1	7	3	33
Intrinsic	3	14	2	13	1	14	1	7	2	22
Tenure	3	14	3	20	0	0	0	0	3	33

b. Enhancing quality

Table 6.3 summarizes the responses of geography department chairs when asked to describe how teaching evaluations are used to enhance teaching quality. All chairs

based their comments on the results of course evaluations. One chair summarized the tone of responses that was received by all chairs:

> "There's no formal method. It certainly depends on the particular chair involved and the initiative of the individual faculty member to try and improve or do better."

Chairs also consistently described a wide range of activities or approaches that could be used to enhance teaching quality. Most chairs described individual meetings that they would have with the individual that was struggling with their teaching and a variety of alternatives that they would present.

> "Look, you're struggling with teaching. There are all sorts of activities on campus to try to enhance teaching quality. There are seminars and such run through the [Teaching and Learning Centre.] They can request to have a teaching mentor come and sit in and observe their class, make recommendations on how to improve, or how to improve examinations, or how to write more effective multiple choice tests, and so on. There are all sorts of literature that will be given to them in terms of the effectiveness of teaching, and current teaching philosophies, and so on."

Cohen (1980) describes the importance of chairs or teaching and learning centre staff and instructors debriefing the data from course evaluations to best determine an implementation plan to effectively use the feedback data. A particularly pragmatic chair described the personal nature of teaching and that enhancements to teaching would likely not change the person's innate ability to teach, but rather enhance their ability to better deliver the course content.

> "In my view, having been in this business for 35 years, is that teaching is a very personal thing. Some people have it, some people don't have it. You can't change the character of a teacher that much. But you can change the way in which they deliver the goods, so to speak."

The majority of chairs (73%) identified the number one mechanism that would be used to enhance teaching quality would be to encourage the struggling instructor to explore the

range of services offered by the Teaching and Learning Centre. A number of these chairs appeared to view the Teaching and Learning Centres as remedial centres, whereas others described the Centres as places that instructors could 'self-avail'. A number of chairs identified that the most frequent users of Teaching and Learning Centres came from two groups, both of whom were trying to enhance their teaching abilities. The two groups were existing good instructors and relatively new instructors. Chairs of research-intensive departments were more likely than chairs of undergraduate-intensive institutions to view teaching and learning centres as remedial centres (80% vs. 57% respectively).

Table 6.3: Mechanisms to enhance teaching quality identified by chairs of Canadian geography departments

Enhancement	Overall		Research		Undergraduate		Unionized		Non-unionized	
	# of Chairs	%	# of Chairs	%	# of Chairs	%	# of Chairs	%	# of Chairs	%
Encouragement for remedial help from T&L Centre	16	73	12	80	4	57	9	64	7	78
Mentor meeting chair/dean	16	73	15	100	1	14	7	50	9	100
Mentor	10	46	9	60	1	14	5	36	5	56
Dismissal	3	14	1	7	2	29	0	14	1	11
Reflection by Individual	3	14	1	7	2	29	3	21	0	0
Development of a teaching culture	2	9	2	13	0	0	1	7	1	11
Development of literature awareness	2	9	2	13	0	0	1	7	1	11

Overall, it appears that the strategies to enhance teaching quality are being applied in a non-systematic fashion, but in a way that may address the specific needs of the individual. It also appears that the assumption underlying this approach is that teaching improvement will happen with practice and without intervention.

6.4 Effectiveness of Current Evaluation Practices

Chairs of geography departments are excellent at critiquing systems and processes. Nearly without fail (>90%) when asked what part of the teaching evaluation practice works best they described in detail what did not work well. A maximum of two additional probes was used to elicit their perspectives on what worked best. Four chairs, three from research-intensive universities and one from an undergraduate-intensive university, did not provide any information about what worked well in the existing system. All chairs were able to provide comment about what did not work well in the existing system. Two possible explanations for this inability to easily address what is working well in the system may be that the system is not working well at all or that chairs are more frequently asked to be critical and have more practice at identifying weaknesses rather than strengths.

a. What part works best?

The vast majority of the comments about what worked best centred on course evaluations. Fifty-five percent of the chairs overall identified the course evaluation tool as the best aspect of the teaching evaluation practices. This result was identical for research-intensive and undergraduate-intensive institutions. There was wide variation in the aspects of the tool identified as working well, with little consistency between chairs. These included the:

- ability to add questions;
- opportunity for open-ended questions;
- regularity at which data are collected;
- quantifiability of the results;
- opportunity to elicit the student voice; and,
- opportunity and regularity of feedback to the instructor.

One of the chairs of a research-intensive university highlighted their observations about why the course evaluation tool is the best part of the process; primarily because it provides either a reward or incentive to instructors about the quality of their teaching. If the instructor is teaching well, the reward comes through the student validation of their work or, if the instructor is not teaching well, it provides an incentive and sometimes ideas of how to improve.

> "I think the student evaluation of educational quality is, probably, the most useful to the instructor. If the instructor is doing a good job it's a nice pat on the back, it's reassurance. If you're doing a poor job, in many cases what I've seen is the instructor wants to improve. It can get them stimulated to do something like that. At the departmental level, it enables the department head to provide advice when necessary. Similarly at the faculty level, I think the same can be said. And I think it's a good thing for students. It is anonymous. It's very confidential. They feel no inhibitions to stating their case. If they really didn't like an instructor, didn't like a course they can do it in the privacy and the confidentiality of the material that's provided back to that instructor. So the grades are filed and about a month later – to ensure students feel that they're not going to be impacted by this – the instructor would receive the bundled package of information, after it's come through the department head, of students' individual comments, with all the bubble sheets, as well as the summary sheets."

In the increasingly accountable university environment the course evaluations were seen as an important tool to some chairs when they are considering their administrative responsibility. The evaluations provided these chairs with information about what was happening in the classroom and how that might impact future undergraduate enrolment. The ideas associated with retention and fiscal restraints are likely to become more common rather than less common in the future, given current budget constraints at most universities within Canada.

> "I think what works best is the information about 'would you recommend that somebody else take this course.' If we've got people who get very low evaluations on those kinds of question we, in the department, are quite concerned

> about that. Our budget is in large part determined by how many bums on seats we have, right? If people are effectively discouraging students to sign up, we'd like to know what we can do to do something about that."

Several chairs identified benefits to the overall system for evaluating teaching effectiveness. One chair suggested that the presence of a systematic approach is key to gathering useful information and ensuring that there is a level of fairness in the system.

> "It is important to have a systematic process. I would argue that works best rather than relying on things like "Ratemyprof" where on a highly electronically literate campus and cohort things can accelerate fairly quickly. I much prefer a systematic approach than that sort of voluntary and less formally structured approach."

The overall system also helped institutions establish their culture around teaching and provided an opportunity to demonstrate behaviours that endorsed that culture. In the discussion with one chair at a research-intensive university, the opportunity to reward teaching excellence through merit pay was an opportunity to validate the importance of high-quality undergraduate teaching and provide a signal on its importance along side the importance of high-quality research.

> "I sometimes think of the whole merit process as something that reminds me of Garrison Keillor's Lake Wobegon where he said that everyone believes that their kids are above average. Everybody expects a merit increase every year in a department like this one where everyone's performing at a very high level. The fact that some of those merit increases go to people who are putting more emphasis and getting more success in teaching than in research, I think is an important signal. It doesn't, necessarily, make those who think that their research has been good but their teaching has been a little off par feel any better about the decision. But that's part of the challenge of having relatively few awards to hand out to people."

b. What part works least well?

The response of geography department chairs to what did not work well in the teaching evaluation process was robust. All chairs were able to identify some weakness in the existing system. The majority of comments centred on concerns about course evaluations, although a number of chairs also identified concerns, primarily related to omissions in the existing system, in the overall evaluative process. Sixty-four percent of chairs identified shortcomings in the existing course evaluation tool. This broke down into 60% of chairs at research-intensive institutions, 71% of chairs at undergraduate-intensive institutions, and 64% of chairs in departments where the faculty are unionized and 56% of chairs in departments where the faculty are non-unionized. The shortcomings that were identified were in many cases specific to the institution, but may have wider general applicability. These shortcomings included:

- inability to evaluate the diversity of learning opportunities in geography (e.g., labs and field experiences);
- length of class time required to complete the evaluations;
- low response rates and concerns about validity of results, particularly at institutions where the evaluations are completed on-line;
- the summative nature of the forms and the lack of opportunity for open ended comments;
- the lack of an effective debriefing of the results with instructors;
- the potential for the instructor to manipulate the results; and,
- the ability of students to provide mature, informed feedback.

The omissions in the existing system that were identified by chairs include:

- opportunity for debriefing with the instructor after course evaluation feedback is provided;
- ineffective instructor feedback;
- lack of meeting the instructor needs of providing constructive feedback;
- lack of objectivity;
- lack of peer evaluation;
- lack of diverse tools beyond course evaluation;

- lack of professional development opportunities related to teaching and learning; and,
- single universally applied tool.

Three chairs explicitly described the teaching evaluation process as being controlled by the instructor's union. In one case, the chair felt that the outcome of the tool used at their institution was to minimize the culpability of individuals.

> "It's [the course evaluation tool] the result of collective bargaining. It's not really there to evaluate teaching so much as to deal with – we have to have an evaluation for administration's perspective. So from the union, what is the least damaging? We end up with questions on the evaluation like: Does your instructor speak clearly and audibly? Is your instructor punctual? Obviously if somebody's not punctual and doesn't speak clearly and audibly they're not going to be a really effective teacher. I think it's measuring very superficial things. It's not delving into what makes somebody a good teacher. From the collective bargaining process, the punitive effects of a bad evaluation are limited by this instrument."

Several chairs (35%) described field experiences as a key component of student learning in geography. The field was seen as an important tool used by geographers for recruitment and retention of undergraduate students. It was also seen as the venue where students had the opportunity to actually 'see' the applied nature of their discipline. The course evaluation instrument was seen to have distinct shortcomings when evaluating field learning, both in the type of questions asked (e.g., classroom temperature) and in the timing of the evaluation. Field teaching was perceived by some chairs as a form of teaching that the value of the experience was often not recognized until well after the end of the experience and beyond the time at which course evaluations are completed.

Few departments that participated in the research conducted their course evaluations through an on-line format (<10%). The remaining departments had students complete the forms as part of class time towards the end of the course. A number of

chairs commented that this is a fairly time-consuming task in the classroom that occurs at a time in the academic year when many instructors are most protective of class time for completing course objectives. If in a typical class it requires 20 minutes to complete these evaluations and a typical student is completing five of these evaluations in a term; with each course having three hours of class time per week over a twelve week term the time it takes to complete evaluations is less than 1% of the class time (i.e., 0.9%).

Several chairs expressed concerns about whether the breadth of student voice was being captured in course evaluations. This appears to be a much larger concern in departments where the course evaluation is completed on-line.

> [With in-class evaluations, in small classes we typically got] "a response rate of 90-100%. That would drop down in large classes. Probably for a large class you would still get a response rate of 60%. The comments weren't very useful; there weren't that many comments. But at least you'd get a large proportion of the students in the course that actually filled in the evaluation. Now that it's on-line one finds a consistent drop of about 20% in the response rate. Then the issue becomes: how typical was that evaluation in class where you get a response rate of 90% down to 60% and what is the typicality of the response rate these days? ... Nobody's really found a way, yet, to increase the response rate back to what it was before. At the present time, faculty are worried that there are now biases – that the only people who'd go on-line to do this are those who have a gripe or grudge against the instructor of the course, or the ones who really are extremely happy. You get the polarization of the views. I don't think anybody's shown, specifically, that that's the case but that certainly is the perception – that people go on there that are really unhappy. In our case there seems to be a sort of lowering of some of the evaluations of questions. Of course because people change courses and student numbers change it's very difficult to prove anything. That quantification and the unreasonableness sometimes attached to it is perhaps the least satisfactory part of the evaluation procedure."

Although a second chair who was interviewed acknowledged that although they had observed a substantial drop in response rate when they went to an on-line form, they were observing similar means and standard deviations on specific questions (e.g., effectiveness

of instructor relative to other instructors). A third institution that had introduced on-line forms reported response rates of less than 10%. With such a small response rate there are obvious, legitimate concerns about the representativeness of the results.

The dichotomy between summative and formative evaluation of teaching was not lost on the geography chairs. One chair commented on the nearly exclusive summative nature of the forms and that feedback did not come early enough in the course to permit change. Educational developers typically respond to that concern that, although changes can not be made in that iteration of the course, the feedback is valuable for subsequent iterations. Educational developers also typically encourage instructors to solicit informal feedback at the midpoint of the course. A second problem with the summative nature of the forms was identified by several chairs and one geography instructor that participated in the research that many course evaluation forms contain a number of summative questions to which neither the instructor nor likely the chair, would have the ability to correct (e.g., temperature of the room, seating in the room). Departments were fairly evenly spaced in the opportunity they provided for instructors to add questions to the evaluation tool in order to solicit open-ended feedback. The open-ended feedback was acknowledged in general as providing the instructor formative feedback.

The process for providing feedback to instructors was widely varied and seen by some chairs as a weakness in the system. At some institutions, the chair met with every instructor at the end of the academic term (or year in some cases) to provide oral feedback about the instructor's feedback. This discussion appeared to focus in large part on the course evaluation results, but may also include student concerns raised in other ways and other evidence of teaching commitment (e.g., workshop participation). At

other institutions, the results are discussed with the chair at the time of the annual review when research and service are also being evaluated and feedback provided. At some institutions the instructors received the results of the course evaluations with no feedback from the chair. Finally, at some institutions, the chair would only meet one-to-one with instructors who, based on the evidence from the course evaluations, were struggling with their teaching. More thorough review and feedback appeared to happen at nearly all institutions during the tenure and promotion process.

Although chairs were hesitant to provide specific examples of instructors manipulating the results of course evaluations, 18% raised it as a possibility and supported by the literature (Feldman, 1979; Braskamp et al., 1984; Marsh, 1984; and Marsh 1987). There were three ways identified in which this manipulation could occur. The first way was through direct conversation with the students. The second involved the timing of providing feedback on student assessment to sway student opinion.

> "I think one of the issues that always haunts the quantitative questionnaire process is the extent to which it's, to some extent, open to manipulation. Parallel to that is the extent to which it is a measure not so much of teaching effectiveness as of popularity. The manipulation issue – what I mean by that is: people who are anxious about their teaching evaluations can time the administration of the questionnaire so that it might follow the return of some grading that has been, shall we say, generous. And then bring the marks back into line with the accepted norms by being draconian on, say, a final exam after the evaluation has been done. There are rules about the administration of these instruments – an outline of protocol and emphasized neutrality in approaching them – but, I think, there's always the opportunity for people to subvert that by saying in the lecture before they administer the questionnaire (when they might, in fact, follow the protocol scrupulously) just happen to offer an aside about their future and the lives of their wives and children, husbands and children depend on their getting successful scores and so on. Putting a guilt trip on the student. I'm not suggesting that that has happened, but that's one of the issues, I think, with the quantitative evaluations. I think there are a variety of ways in which they can be suspect in that regard."

The third way in which course evaluations were seen as being prey to manipulation was related to workload and difficulty of course material. The following discussion between the interviewer and a geography department chair highlighted this concern.

> "I think student bias could, particularly in smaller classes, skew the results. I'll leave it at that."

Ok. I'm going to ask a follow up on that. Do you think that there's a possibility in a smaller class setting to become a popularity contest? Is that where you're going with that comment?

> "Yeah, I think so. You have certain questions like, "Was the workload sufficient for this course?" Any professor that has a heavy workload, this happens to me all the time, you have a heavy workload because I feel students learn more when they have to work more. But students don't necessarily see it that way. So you can get nailed on questions like that. "No, there was way too much work." But then you get a question like, "Did you learn a lot in this course?" And you score high on that one because they have a high work load. There are conundrums like that that are fairly obvious. Yeah, it can become a popularity contest."

Nearly 35% of the chairs in both research-intensive and undergraduate-intensive departments raised some concerns about student ability to effectively evaluate teaching.

> "In terms of what they measure, I think this is a much more serious issue. We really don't have good information on, as it were, the psychology of students responding to these evaluations. They, obviously, get deluged with them when every course is being evaluated by mandate. They will have at least five of these things to fill out in the last week or so of term, usually. What the impacts of that are on the kind of accumulated repeat behaviour is not something that we know a whole lot about. There is always concern about how learning takes place and what can be evaluated. Some of the learning that is done there may not be appreciated until way, way later. So this can't be evaluated. And some of the questions on the questionnaires are also ones that perhaps incline towards an evaluation of the mechanics and techniques of teaching and student preferences for certain styles not necessarily related to intellectual challenge."

Another chair raised the issue of equity or fairness of the students in evaluating teaching.

> "We follow in our department, of course, what the university says we should. But I'm not entirely sure that it's completely objective. There doesn't seem to be any way, really, for student evaluation in particular. In 90% of the cases, I have got to assume students are fair, but there's no guarantee that they are. There's no guarantee that the grading that they assign are based on objective or subjective views. For example, a student who feels the course is too hard may describe the instructor a jerk type deal. I guess that's my personal feeling. I know we have to have an evaluation process of some kind. There has been enormous debate in our institution over whether or not this is the best way to do it. But this is the way they do it and we sort of have to live with it."

The concern of chairs about students' ability to objectively evaluate and the concerns that instructors may be able to manipulate the results are confirmed in the literature.

> "When grades and exam scores are significantly correlated, then higher evaluations by students may be due to (a) more effective teaching that produces greater learning and higher evaluations by students; (b) increased student satisfaction with higher grades which causes them to 'reward' the instructor with higher ratings independent of more effective teaching or greater learning; or (c) initial differences in student characteristics that affect both teaching effectiveness and performance." (Marsh, 1987 p. 290)

6.5 Model of an effective teaching evaluation system

As Alice is often paraphrased to have stated in *Alice in Wonderland,* 'if you don't know where you are going, any road will take you there', succinctly describes some teaching evaluation systems in higher education (Carroll, 1866). The challenge observed in this research with many teaching evaluation systems is a lack of clarity about the goals for teaching evaluations and the process for evaluating teaching. In this chapter a systematic teaching evaluation model (i.e., a system for evaluating teaching) will be described.

The stakeholders in the teaching evaluation system are described in Section 3.1 (see Figure 3.1). The internal stakeholders include: students, instructors, instructional

staff, departmental chair, faculty (i.e. dean), Faculty Association, and institution. External stakeholders include: society, parents, employers, and provincial and federal governments. This wide range of stakeholders creates a need for an effective balance to be obtained between the wide-ranging views. A chair at a research-intensive university summarized well this need for balance between the needs of quality control (i.e., accountability), the needs of instructors for reflection on teaching and the needs of the department to enhance programs and curriculum.

> "I do think that we should be trying to look at ways in which we can reconcile the institution's need to be seen to be concerned about quality control with instructors' desires to inform themselves better about the way in which their teaching is received and about how effective it is so that we can develop programs that serve both those goals. I really don't think that we've got that at the moment. I think the balance is much too heavily towards the institution's concern with quality control rather than really focusing on educational issues. I think, actually, one of the things that's really lacking here – and probably more generally in North America – is the lack of external moderation of exams and things like that to provide some sort of insight into levels of consistency between the courses within the institution and between institutions in terms of expectations and educational outcome."

Using Lewin's force-field analysis approach of identifying the internal and external factors driving and restraining the development of a system provides a framework to identify the often competing demands and distinct purposes for evaluating teaching, as well as highlighting the challenges of creating an ideal single system (Lewin, 1951). The external forces, including the nature of the discipline, the expectations of society, the needs of government, economic reality, peer institutions and parental expectations all impact how teaching is evaluated within higher education. The internal forces, many of which are evident in the demographics and culture of place, also impact how teaching is evaluated. Both are discussed with respect to their driving and

restraining of change more fully in the remainder of this section. One chair succinctly described the dilemma that results from the dynamic nature of the student population and how it impacts teaching.

> "One of the difficulties is the interest of students and the background of students is changing more rapidly than we think it is. One of the problems that we face, I think this is not just geography but probably other disciplines as well, is that our teaching styles and the way in which they're being received by the students is changing at a fairly rapid rate both in terms of their ability to be open to electronic media that depends upon WebCT, PowerPoint, and all the rest of it. I'm not sure how we are going to be able to adjust to these rapidly changing student bodies."

To better understand how these forces impact a teaching evaluation system, a schematic model of an effective teaching evaluation system is displayed in Figure 6.1.

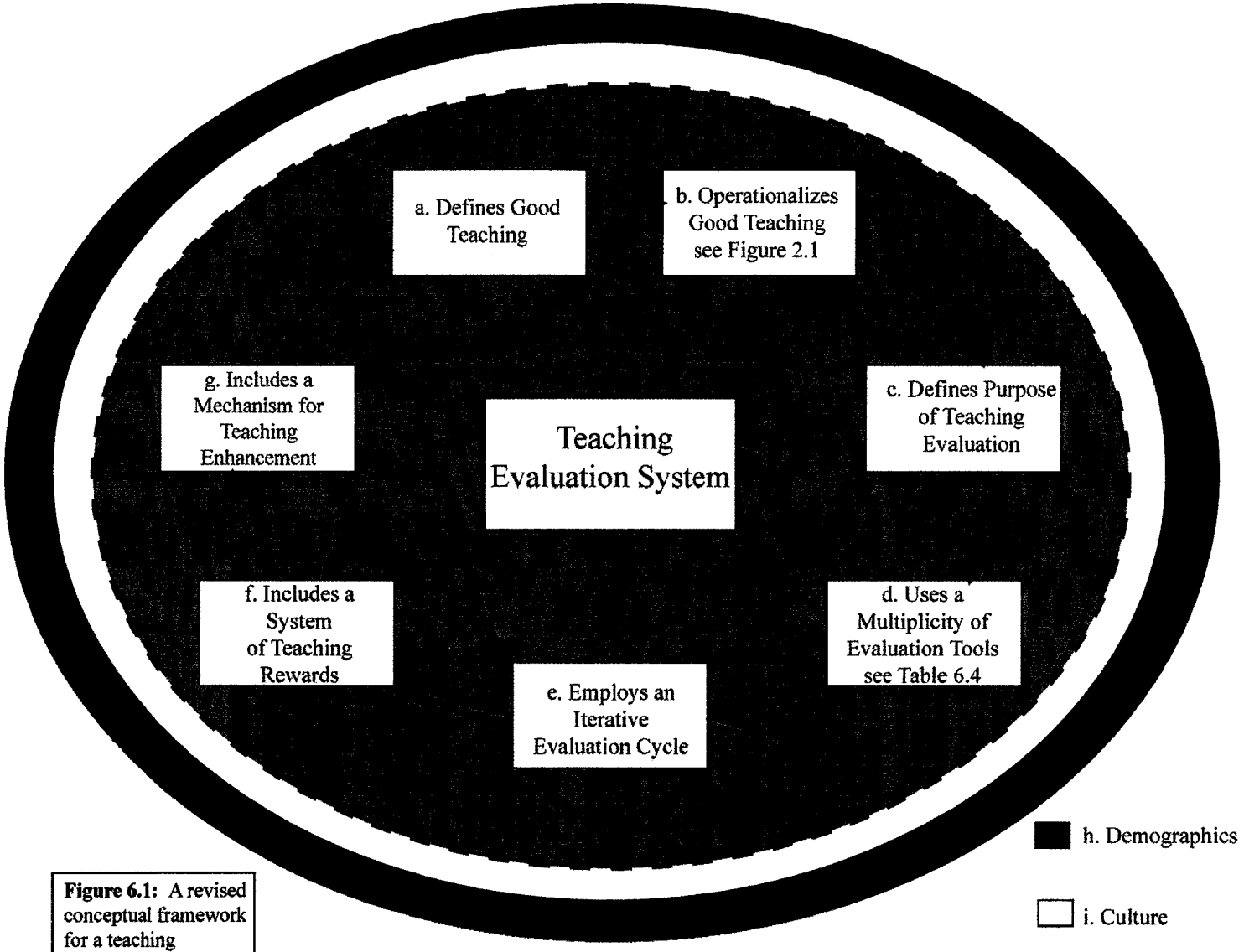


Figure 6.1: A revised conceptual framework for a teaching evaluation system.

a. Definition of good teaching

A good teaching evaluation system must clearly define and describe the breadth of what is considered good teaching (Figure 6.1). The definition of good teaching needs to be public for stakeholders to access, particularly those that are internal to the process (e.g., students, instructors, administration). The definition of good teaching should include the meta-characteristics identified by geography chairs in Table 5.1. The meta-characteristics include: student engagement, knowledge transfer and student learning. The literature suggests that an additional meta-characteristic, skill development, should also be included (Young and Shaw, 1999; Feldman, 1976; Braskamp et al., 1984). A possible definition of good teaching that would incorporate the four meta-characteristics is:

> Good teaching is teaching that results in student learning (e.g., change in attitudes and understanding) through student engagement, and results in high levels of student learning, including knowledge transfer and skill development.

The definition of good teaching is universal, but the weights assigned to the individual meta-characteristics may be dependent on the culture of the institution. For example, in Table 5.1 it appears that research-intensive universities place a heavier weight on knowledge transfer whereas the undergraduate-intensive universities tend to place a greater emphasis on skill development.

b. Operationalizing good teaching

The review of literature was used by the researcher to create the conceptual framework described in Figure 2.1 which summarized the attributes of effective teaching. Table 5.1 demonstrated that the attributes of effective teaching identified in Figure 2.1

corresponded quite well to the good teaching characteristics identified by geography department chairs. Instructors would benefit from an awareness of the literature associated with effective teaching. A moderately easy way to ensure instructor awareness of this literature is by communicating the characteristics of effective teaching as described in Figure 2.1. This communication could occur as part of the feedback discussion (described more fully below) subsequent to teaching evaluations.

Geography chairs were fairly consistent in their identification of additional attributes of good teaching that were discipline-specific including: experiential learning, visual presentation and interpretation of data, the global and interdisciplinary nature of the discipline and the need for technological currency. It seems reasonable that other disciplines would also identify several attributes that were specific to their discipline. As part of ensuring credibility of the teaching evaluation system it is important that each discipline within an institution identifies and communicates with internal stakeholders those items that are discipline-specific.

c. Purpose of teaching evaluation

The purpose for evaluating teaching needs to be clear for both evaluators and those being evaluated (Figure 6.1). This research has demonstrated four key teaching evaluation purposes: department and institution accountability (i.e., quality control) (Murray, 1980; Marsh, 1987; Marsh and Roche, 1997); instructor feedback (Murray, 1980; Marsh, 1987; Marsh and Roche, 1997); administrative decisions (e.g., hiring, tenure, promotion, merit increases) (Murray, 1980; Marsh, 1987); and, capturing the student voice (Murray, 1980; Marsh, 1987). Two other purposes, which are described in

the literature and appear to be reasonable purposes for evaluation, curriculum and program development (Marsh, 1987); and, research on teaching and learning (Murray, 1980; Marsh, 1987) were not identified in this research by chairs of geography departments. The tools used to evaluate for the different purposes do not need to be mutually exclusive. For example, the course evaluation instrument may provide an instructor formative feedback about teaching and include several summative comments that contribute to institution accountability, administrative decisions (e.g., tenure, promotion, merit increases). In the following section different teaching evaluation tools and the linkages to the different purposes will be described

d. Multiple teaching evaluation tools

A good teaching evaluation system will use multiple tools to measure teaching effectiveness (Table 6.1). This is paramount to ensure that triangulation amongst: the breadth of what constitutes good teaching is being evaluated; the needs of a wide range of stakeholders in the teaching evaluation system are being met; and, the purpose(s) of evaluating teaching are being addressed. These tools must meet a number of criteria including being valid, reliable, credible, relevant, impartial, and statistically sound (see Section 3.3). Each of the tools described in Table 6.4 will be described in more detail below.

Table 6.4: The purposes of teaching evaluation, the type of evaluation and the possible tools that can be used to evaluate (*the purposes for evaluating teaching are described in Section 3.2; [+]there are a number of other tools that can be used in curriculum and program development including program dossiers and student learning outcomes, this table only describes those related to teaching evaluation)

Purpose*	**Type of Evaluation**	**Tool**
Department & Institution Accountability	Summative *	Course Evaluation Teaching Dossier Annual Report
Instructor Feedback	Formative	Course Evaluation Teaching Dossier Annual Report Self Reflection Peer Evaluation of Instruction Content of Course Materials
Administrative Decisions - hiring	Summative	Course Evaluation Teaching Dossier Content of Course Materials Mock Presentation Student Focus Groups Round Table Discussion with Faculty
Administrative Decisions - tenure	Summative	Course Evaluation Teaching Dossier Content of Course Materials Student Focus Groups Alumni Feedback
Administrative Decisions - promotion	Summative	Course Evaluation Teaching Dossier Content of Course Materials Student Focus Groups Alumni Feedback
Administrative Decisions - merit	Summative*	Course Evaluation Annual Report
Student Voice	Formative & Summative	Course Evaluation Student Focus Groups for hiring & for tenure and promotion
Curriculum and Program Development	Formative & Summative	Course Evaluation[+]
Research on Teaching & Learning	Formative & Summative	Course Evaluation

Course evaluations are used universally across Canada within geography departments and their use is well-supported in the education literature (see Chapter 3). They have been shown to be reliable and valid. There are a number of commercially available forms which are used at a couple of institutions within Canada, although, most institutions, Faculties or departments have designed their own forms.

To ensure impartiality in the use of course evaluations, a system must be in place that ensures they are not being implemented in a way to introduce bias (e.g., immediately following the return of student assessments with particularly high grades or following an introduction by an instructor that could influence students to be more generous in their evaluation). Ideally, the instructor should be outside of the classroom when evaluations are completed and the evaluation process should be introduced by someone other than the course instructor (e.g., departmental support staff, student, another instructor). To ensure common instructions are received by the students, a script, overhead with instructions or short PowerPoint should be used. The instructions should include the definition of good teaching that is being used.

Interviews with research participants indicated a wide variation in the level of credibility of different course evaluations. This may be due, in part, to the culture of the institution. Instructors need to be provided evidence of why different questions are being asked. It is reasonable to expect that an instructor may wonder about asking questions concerning room temperature and seating characteristics on an evaluation tool for which the data will not be shared with the institution's physical plant in order that they may actually be able to make improvements in this area. As well, instructors need to clearly understand which portion of the evaluation will be used for summative purposes and

which for formative purposes. At some institutions, a portion of the course evaluation data is made available for summative administrative purposes to deans and chairs and other portions of the data are only available to instructors for formative purposes. For example, questions about the instructor's abilities relative to colleagues or overall effectiveness are often used for strictly summative purposes, whereas questions about the workload, perceived level of difficulty of the course content and delivery style are often intended strictly for formative purposes. At one research-intensive university a chair questioned which questions were being used for summative and which for formative purposes. For example, the chair questioned the inclusion of the instructor's promptness at lecture being included in the formative component of the evaluation and not in the summative portion. The chair argued that if an instructor is continually late for class, administration needs to be aware of the problem in order that they may work to correct the problem. The introduction of a short handbook that communicates to the instructors why the evaluations are being completed (i.e., the purpose), the mechanism for the development of the form, and suggestions on how to use the information gathered from the form would enhance the feedback cycle and increase the credibility of the tool being used.

There appeared to be a particular lack of credibility in the course evaluation tool within departments whose instructors are unionized. This may, in part, be a function of who is responsible for the creation of the tool. Summative evaluation tools that are approved by unions appear at some institutions to ask questions and gather data in ways to ensure that an individual union member is not personally responsible for poor teaching. It seems reasonable that formative evaluation tools could be effectively used in a

unionized setting if the data were only available to instructors and not to the administration. Perhaps in a system where the administration pays the costs associated with collection of the data but the union provides the staff required for the evaluation and distributes the results of the evaluations, would address the concerns and needs of the different groups.

Teaching dossiers have been widely accepted within Canadian universities as a mechanism to document teaching. Different institutions have different suggestions for materials to include in the dossier which may, in part, depend on the culture of the institution. In general dossiers include: a teaching statement (or philosophy), data from course evaluations, list of courses taught, solicited or unsolicited comments from students, statements from colleagues, and documentation of teaching innovation or use of new teaching strategies (Seldin, 1991). The act of preparing and reviewing the dossier should provide the instructor some formative feedback on their teaching. The dossier is typically used by administration to make summative decisions about hiring, tenure, and promotion. Chairs of geography departments reported wide variation, on the usefulness of the teaching dossier for hiring. Some departments required the dossier, but would only sometimes receive it from potential candidates. Other departments reported that dossiers were not very useful tools in the hiring process because candidate often had very little teaching experience to document.

An ***annual report*** is mandated in nearly all departments of geography within Canada. The annual report is prepared by each faculty member and typically summarizes their teaching and research activities of the past year. The annual report is used extensively as evidence for merit increases. Typically, the teaching component of the

annual report includes a summary of courses taught, theses supervised (undergraduate and graduate), and course evaluation data from one or more of the summative questions. Two chairs described an enhancement to the reporting of the summative course evaluations. These chairs reported using rolling averages to report summative course evaluation data. This means that the numbers reported would be based on three years of instruction by the individual in the course. If an instructor had taught the course for less than three years, the course would be noted, but no summative data reported. This ensures that instructors are not being penalized in a merit scheme during the first year or two that they teach a course when typically course evaluation scores are lower. Only one research participant described a formal part of the annual report that permitted faculty to document teaching innovation. Teaching innovation in this case meant implementing teaching strategies that were new to the individual. Two chairs commented that, although faculty at their institution complete annual reviews and merit increases are based on the annual review, nearly 100% of the faculty receive full merit increases each year, despite there being a wide range in the content of the reviews. Thus, the annual reviews are not being effectively used to discriminate differences in merit pay. The problem of effective discrimination in determining merit pay associated with teaching excellence was described equally by chairs/instructors at unionized, non-unionized, research-intensive, and undergraduate-intensive institutions. This suggests that additional measures of teaching quality and/or training in effective evaluation or interpretation of teaching quality data may need to occur.

At the heart of research is the peer-review process whereby one receives feedback, evaluates the feedback, and responds to the feedback. As a result, academics

frequently engage in reflection as part of the peer-review process in research. The opportunity for reflection is not formalized to nearly the same extent in teaching. Figure 6.2 describes an iterative cycle of feedback and reflection that an ideal teaching evaluation scheme would contain. Ideally, part of the reflection stage would be ***self-reflection***. In research it is accepted that there will be written documentation of the reflection (e.g., response to reviewers comments). To ensure credibility of the same process, it seems reasonable to ensure a written component in teaching self-reflection. This could occur in the form of a short written statement contained within Annual Reports that identified areas of strength and areas of weakness within an individuals teaching over the past year. Some people may argue that to be most effective the self-reflection should be formative in nature and not summative. It is true that the annual report is a summative document, but if parallelism to research is desired one must remember that ultimately the peer-review process is a summative process (i.e., the reviewers comments determine if the paper will or will not be published). Pinsky and Irby (1997 p. 976) concluded from a survey of successful clinical teachers that,

> "Learning to teach involves a process of turning instructional failures into improved teaching."

Chairs described a system of ***peer evaluation*** that was currently under-utilized and often an inappropriately utilized component of most teaching evaluation systems in geography. Peer evaluation is currently used most commonly for administrative decisions regarding tenure and promotion. It typically involves a colleague observing one or two classes, browsing instructors' course evaluations and perhaps informally chatting to students. Research participants who reported having been peer evaluators commented that they did not feel that the system worked well. They were unprepared to give particularly critical feedback that would be used for summative purposes (i.e., promotion or tenure) based on minimal data, no formalized training in how to evaluate and recognition that they may be a colleague of the individual being evaluated for the next twenty years. Peer evaluation for summative purposes is used in research, as described above, for the peer -eview process and as part of the tenure and promotion system. For the tenure and promotion process it is considered necessary to obtain peer review from individuals external to the department to ensure credibility of the candidate's evaluation as they proceed through the faculty and institution stages of the process. A similar 'arm's length' review of a candidate's teaching is usually not obtained. This is a flaw in most systems. Increasingly there are discipline experts in pedagogy and there are educational developers who would be quite capable of rigorously evaluating these candidates. This 'expert' peer evaluation would ideally involve, an evaluation of a candidate's teaching materials along with classroom performance, whereas the current system rarely reviews the teaching materials (e.g., assessments, course outlines, and

classroom activities). Peer evaluation by a departmental colleague would be an effective teaching evaluation tool if used for a formative purpose. Using it for a formative purpose would relieve the evaluator's concerns about collegiality with a colleague and minimize concerns a bit about the quantity of material available for use in evaluating. To ensure the parallelism with the research, a peer-review that evaluated a paper product should occur. In research, the paper product is the paper and in teaching the paper product would be the course materials.

As described above, the ***review of course materials*** could be done by an 'expert' (e.g., discipline colleague at another institution or an educational developer) for summative purposes or by a departmental colleague for formative purposes. In a department with a high level of collegiality and a supportive teaching environment (see culture of place section below), regular review of course materials by peers would occur.

Ensuring that candidates that demonstrate good teaching potential are hired was described by three chairs as the most effective strategy that could be used to ensure that teaching quality was high within the departments. One common element to more than 90% of the departments was the use of a presentation to evaluate teaching during the interview process. The evidence from this research indicates that ***mock presentations*** are the best way to gauge a candidate's ability to present information effectively for teaching. As one chair stated, ensuring that the candidate showed good promise as an effective teacher was the best that you should expect in an interview. It was unlikely that a potential candidate, particularly for an assistant professor position, would have extensive evidence of teaching excellence. There appeared to be some discrepancy about what would be the ideal characteristics of good teaching potential. Some chairs linked

teaching potential to evidence of the attributes of effective teaching which are described in Figure 2.1. Other chairs suggested that, although you would like to see some of those attributes in the mock presentation, you were also interested in linking the mock presentation style to comments and behaviours that the candidate exhibited in other parts of the interview process; particularly, comments and behaviours that related directly to undergraduate students. Ultimately, both groups of chairs are suggesting that the ideal candidate to hire should demonstrate potential for teaching excellence which can be triangulated from attributes of good teaching observed in a presentation, and attitudes and behaviours observed throughout the interview process.

Student focus groups provide another mechanism beyond course evaluations to capture the student voice. Although focus groups take time to facilitate and dissect the results from they can be used for both formative and summative teaching evaluation. Student focus groups can provide instructors ample insight into areas for improvement. Repeated focus groups, using a panel design, would allow instructors to gauge improvement over time. For summative purposes, student focus groups can provide primarily qualitative data that can be used for promotion and tenure decisions. Ideally, the focus group would provide a deeper level of understanding of areas of strengths and weaknesses in teaching effectiveness than course evaluations provide.

Round-table discussions which are attended by a number of faculty and the potential candidate for hire have proven very effective at two research-intensive universities. The format for these discussions is a 60-90 minute period in which the candidate is asked questions in a slightly less formal way than other components of the interview process by their potential departmental colleagues. These questions cover the

spectrum of research, teaching and service. They are quite broad and may relate to existing experience or ask the candidate to comment on future directions. These discussions were reported to very helpful in illuminating for departments if the candidate would be a good fit to the school. Part of the idea of fit would be the fit to the existing curriculum and teaching philosophy within the department. The idea of fit ties quite closely to the sections below on culture of place and demographics of place.

One chair at a research-intensive university expressed a strong opinion that quality of teaching often could not be measured until several years after a course was completed. This idea is alluded to in some of the literature on teaching and learning as it relates to field teaching (Tricart, 1969; Gold et al., 1991; Abler, 1994; Cooke, 1998; Chalkley et al, 2000). The gathering of ***alumni feedback*** on teaching would be a very effective way to ensure that teaching effectiveness beyond the immediacy of the current classroom experience was being assessed.

e. Iterative teaching evaluation cycle

A good teaching evaluation system should be based on an iterative cycle (Figure 6.2) that contains a reflective component at each stage (Carini, 2006). Evaluation must lead to: reflection; the opportunity to receive feedback; which leads to further reflection; and, the opportunity for implementation of ideas based on the feedback; which leads to reflection; and, further opportunities for evaluation. Current systems employed within Canadian geography departments provide ample opportunity for evaluation (often singular evaluation of student perceived effectiveness), but rarely provide significant opportunity for reflection or a mechanism to evaluate implementation of change.

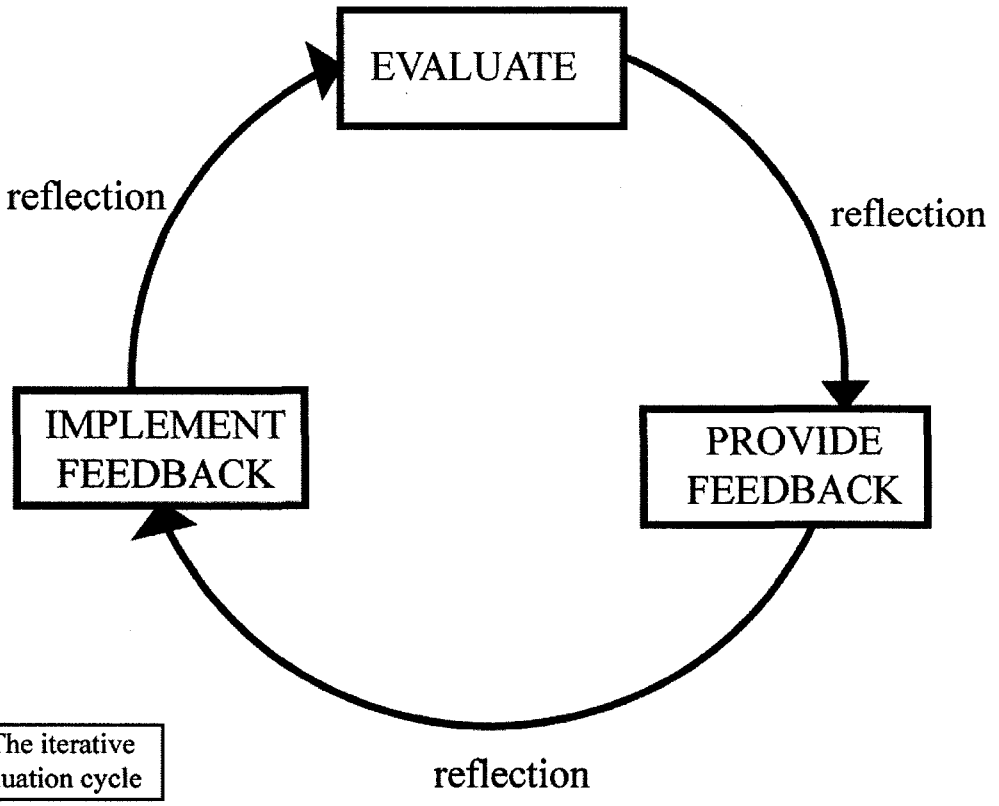


Figure 6.2: The iterative teaching evaluation cycle

f. System of teaching rewards

An effective teaching evaluation system should include a system to reward excellence (Abler, 1994; Boyer, 1990; Murray, 1973). Rewards should be a combination of extrinsic and intrinsic. Extrinsic rewards include a meaningful merit system, an award system at the institution and provincial level, and the opportunity for teaching release to further enhance teaching. The award criteria need to be clearly and effectively communicated to the different stakeholders. The annual review and merit system, according to chairs at both research-intensive and undergraduate-intensive, unionized and non-unionized institutions has struggled due to a lack of transparency and communication surrounding teaching expectations. Parity with the systems in place to evaluate research

(discussed below in section g) including peer review, formative and summative feedback, and clear indicators of success (e.g., a clear indication of research success is a publication, but a clear indication of teaching success is often less well defined) should be sought for teaching reward systems.

The intrinsic rewards, although harder to quantify, may be equally or more important. The intrinsic awards need to be ingrained within the culture of place (see section below). Examples of awards that would be primarily intrinsic in nature would be the opportunity to participate as a mentor or engage in peer review. Individuals that place a high level of importance on teaching excellence would likely view the opportunity to participate in these activities (if they are perceived as valuable to the culture of the department) as rewards. An administrator at a research-intensive university describes the translation of teaching excellence into an opportunity to explore teaching pedagogy as a reward:

> "You can use it as a vehicle for other areas of development. You can certainly encourage pedagogy of teaching within a discipline. You actually can have people who are strong teachers take that into a research mode or take that into a publication mode or help build that piece of their career academically."

Purely intrinsic rewards could also include the feeling of success that an instructor may feel after a student demonstrates learning or the chance to satisfy one's own curiosity or a feeling of usefulness that can result of having an impact on student learning. Although these are difficult rewards to explicitly cultivate, the creation of a teaching and learning environment in which value is placed on teaching should validate some of these intrinsic rewards (Healey, 2000).

g. Mechanism for teaching enhancement

In this research there emerged four main groupings of methods to enhance teaching quality. These include: the use of teaching and learning centres; mentoring, movements for change and special programs. Each of these will be explored in more detail.

This research has demonstrated that there exists on many university campuses disconnect between academic departments and ***teaching and learning centres*** about the most effective way to makes use the centres. Many members of academic departments including chairs and instructors (>78%), described the teaching and learning centre on their campus as being centres for remediation of teaching challenges within their departments. If a department chair, through the evaluation of teaching, identified a shortcoming or weakness in an instructors teaching, they would often describe referring the individual to the teaching and learning centre for assistance. Chairs described this assistance in the form of participation in workshops, one-to-one meetings with teaching and learning centre staff, and mentoring. All staff at the teaching learning centres interviewed for this research, sometimes at the same institutions as the chairs that described the centres as being remedial centres, did not see their centres as remedial. Rather, they described a system whereby faculty would identify a strength, weakness, or question and would solicit the assistance of staff or physical resources available through the centre. Teaching and learning centre staff were unanimous in their description of self-identification of need rather than a referral that could be perceived as punitive as being their operating style. To achieve an effective teaching evaluation system, this

disconnect between user (i.e., chair and instructor) and supplier (i.e., teaching and learning centre) about the perceived purpose, which appeared to be widespread in institutions across the country, might best be addressed in a structural way (e.g., by providing clarity to teaching and learning stakeholders about the role and responsibilities of teaching and learning centres). Teaching and learning centres were also seen by chairs, instructors and teaching and learning centre staff as key players in facilitating awareness of what constitutes good teaching, and how to improve teaching.

Mentoring was seen as an effective tool to enhance teaching quality. In all departments that participated in this research, some form of mentoring programs were established when a tenure-track faculty position commenced. None of the research participants described a mentoring program for sessional or contractual-limited positions. Participants were not directly asked if these programs existed; thus, the lack of mention does not mean that they do not exist, simply that their existence was not described as part of the teaching or teaching evaluation practices. Mentoring was also introduced in response to poor teaching evaluations. Chairs often became the mentor in response to poor performance. No department described a process of mentoring that treated teaching and research distinctly (i.e., a different teaching mentor than research mentor). If the objectives of the mentoring program are to enhance teaching effectiveness and to enhance research effectiveness, it seems reasonable that a single mentor may not necessarily be the best approach although, a dual mentoring relationship may set up the mentee for potential failure if the dual mentors are providing conflicting direction to a new faculty member. All mentoring that was described by chairs occurred internal to the department.

It was surprising that none of the formal mentoring relationships explored matching of partnerships outside of an instructor's discipline.

At a number of institutions chairs described a ***movement of change***. There was a general lack of overall consistency in the definition of that change, but some common characteristics did emerge. There was discussion of a general increasing emphasis being placed on undergraduate education and specifically teaching, particularly at research-intensive universities. This aligns well with the ideas described in Chapter One (The Rae Report, 2005; The Boyer Commission, 1998; Smith, 1991; Boyer, 1990) about a renewed commitment to undergraduate education. There was an underlying current among undergraduate-intensive institutions that the value and expectations being placed on instructors to engage in research was increasing. There appears to be a compression in the range of expectations placed on faculty between research-intensive and undergraduate-intensive universities. Those at research-intensive institutions are describing a commitment to undergraduate student education and those at undergraduate-intensive institutions are describing a commitment to increased research productivity. The breadth between the expectations for teaching versus research appears to be narrowing across the country in geography departments. At this point in time, the researcher could find no direct evidence that this change in expectations had actually been realized, but there was a definite discussion of narrowing of the breadth of expectations. At all institutions, there was a discussion of flux and change. For this change to be realized and sustainable, the emphasis needs to be placed on the understanding and evolving the culture of place (see below) within the confines of the demographics of place.

A limited number of staff at teaching and learning centres were research participants. A web scan of teaching and learning centre web sites across Canada suggests that although all centres provide some similar programming, there is also a wide ***program*** variation. Centres that have the advantage of Faculties of Education at their institutions have sometimes been able to effectively provide synergies between education students and faculty to enhance teaching effectiveness.

h. Demographics

A key parameter of a good teaching evaluation system is that the system must consider the demographics of the institution and the department. The demographics may influence the weight assigned to the different components of the teaching evaluation system (e.g., the tools used to evaluate teaching). The demographics may be influenced by:

- size (e.g. institution, graduate program, undergraduate program, department);
- administrative structure (e.g., reporting structure);
- level of research-intensiveness;
- nature of the student body (e.g., residential versus commuting population, age, gender, disciplines, socio-economic status)
- nature of the instructional body (e.g., tenured versus contract staff, availability of teaching assistants, age, gender)
- availability of teaching resources; and,
- status of the institution-employee relationship (i.e., are teaching staff unionized).

The size of the institution may have a number of impacts on the teaching evaluation system. For example, a smaller institution may have the ability to allow greater flexibility in the evaluation tool and provide a different approach to providing feedback (e.g., meeting with a dean rather than a departmental chair). In a larger institution, there may be a greater need for a global course evaluation tool rather than

allowing individual Faculties or departments the opportunity to design their own tool simply because of the logistical hurdles involved with multiple tools in a large institution. As well, in an institution with a unionized relationship between administration and instructors, there is likely a Collective Agreement which may require a single evaluation tool to ensure consistency of the treatment of its membership.

The administrative structure of an institution contributes to its demographics of place. There is often correlation between the size of the institution and the administrative structure. The administrative structure will determine the reporting structure for activities like course evaluations, and assessing rewards and enhancements for teaching success. As well, the administrative structure will impact the relative weights of the department, faculty, institution and potentially teaching and learning centres to the culture of place.

At the onset of this research, the hypothesis was made that the level of research-intensiveness would negatively correlate to teaching value. In other words, an institution or department that was identified as research-intensive would place a lower value on undergraduate teaching and conversely an institution that was identified as less research-intensive and more teaching intensive would place a higher value on undergraduate teaching. The data did not support this hypothesis. There was not a direct correlation between level of research-intensiveness and value on undergraduate teaching (see Tables 5.1, 5.2 and 5.3). There was a lower teaching load at those institutions that were identified as research-intensive. This lower teaching load appeared to allow, in some cases, more reflection on teaching effectiveness than those institutions with much higher teaching loads. Interestingly, one institution that was identified as being teaching

intensive actually identified that a greater emphasis was placed on research than on teaching merit.

> "In the department itself we look at teaching, research, and service. I think – this is going to be no surprise – the research file is the one that's going to take precedence."

There was an expectation at teaching intensive institutions that teaching of good quality was the expected norm.

The nature of the student body will influence the teaching and learning environment, thus, it will impact the teaching evaluation system. For example, an institution with a greater number of commuting students may have a greater (or at least different) challenge in creating a strong sense of community amongst the student body. As well if an institution is comprised by a large number of 'mature' students this will influence the expectations and perhaps the mechanisms of delivery of teaching and learning. This will in turn impact the teaching evaluation system that would ideally be employed.

The nature of the instructional body will impact the demographics. At an institution with a large number of contract academic staff the needs of these individuals may be different for the type and amount of feedback that they desire about their teaching effectiveness.

The availability of teaching resources will impact the teaching evaluation system. There was great variation in the reported size, significance, influence and history of teaching and learning centres across Canada. There did not appear to be any relationship between level of research-intensiveness and level of teaching and learning centre. No

solid relationship emerged about the size of the teaching and learning centres and the relative size of the institution. Although all chairs were aware that centres existed on their campus, there was fairly wide variation in the level of use made of the centres. The characteristics, including the reporting structure, of the teaching and learning centre will influence the demographics of place which in turn will impact the teaching evaluation system (e.g., resource availability, literature awareness of instructors about teaching and teaching evaluation).

Likely, the number one influence on the demographics of place as it relates to teaching and teaching evaluation is a function of the institution-employee relationship. A unionized versus non-unionized faculty work environment will make a significant impact on the model of teaching evaluation. This is also discussed in the section on teaching evaluation tools. The entire relationship between the employer and the employee appears to change in an unionized work environment. The chair who described the teaching evaluation tool used at their institution, designed and approved by the instructors' union as designed to minimize the culpability of the individual instructor, effectively illustrates the extent to which this impact can occur. From the observations gained through this research project, establishing a culture of place that is conducive to valuing undergraduate teaching and learning may be more challenging to develop, but also more critical to effectively develop in an unionized environment.

i. Culture

A good teaching evaluation system cannot exist in isolation rather it exists within a supportive culture. The culture in this case refers to the psycho-social dynamics of

institutional relationships. That is, the psychology and sociology of the different relationships that exist at a university. The culture is dynamic, it is flowing or evolving and has an associated direction (Hannerz, 1993). Some of the components to the culture would include the mission and vision statements of the institution. Along with the mission and vision would be the current trajectory that the institution is on relative to those statements. In order to ensure the highest likelihood of a teaching evaluation system succeeding, it must be accompanied by a perception among the stakeholders that the department, faculty, institution, government and society value teaching. Ideally, this value must be based on stakeholder observations of both words and behaviours that demonstrate a commitment to undergraduate teaching. The primary determinants of teaching culture are department, faculty, teaching and learning centre, institution, and discipline (Figure 6.3). The chair is seen as being integral in determining the culture within an academic unit with respect to the culture surrounding teaching.

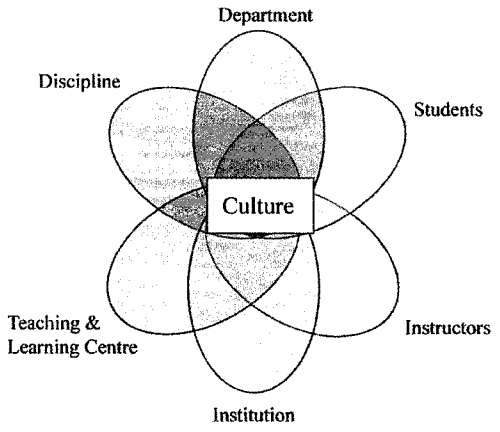


Figure 6.3: The determinants of teaching culture

A good teaching culture within a department will work to minimize the isolation and fragmentation that individuals who choose to focus more resources (time and energy) on teaching rather than research often experience. Often these individuals are working in isolation or within fragmented groups with limited resources. Massy et al., (1994) completed more than 300 interviews with faculty in Science, Social Science and Humanities at 11 colleges and universities (eight research institutions, four doctorate-granting institutions and three liberal arts colleges) about the context of their work. The outcome from this work was that these researchers were able to identify practices that worked against the pressures of fragmentation, isolation and lack of resources to effectively support teaching. The idea of a supportive culture, one where the members value and take teaching seriously, was seen as a foundation. There were a number of attributes that contributed to this foundation including:

- frequent interaction among members to discuss teaching, both formally and informally;
- the ability to tolerate differences of opinion and ensure that they do not lead to divisions within the department;
- generation equality between senior and junior members of the department;
- workload equity (i.e., all faculty teach and all faculty teach all levels);
- course rotation with the expectation of sharing course materials;
- peer evaluation of teaching;
- credibility and validity of course evaluation scheme;
- evidence of balanced incentives (i.e. evidence that teaching is highly valued and critically evaluated for administrative decisions like tenure and promotion)
- the use of consensus decision-making whenever possible;
- an effective chair; and,
- an overall sense of collegiality (Massy et al., 1994).

Peer evaluation of teaching can reasonably be expected to be more widely accepted and effective in a department with a supportive culture where teaching is often

discussed among colleagues. The idea of course rotation may encounter distinct hurdles being implemented in Canadian geography department with distinctly limited resources. Massy et al. (1994) argued that the department chair may be the most important attribute of creating an environment that values undergraduate teaching. What are the characteristics of the effective chair in this regard? The effective chair would need to have a suite of personal skills, including the ability to advocate on behalf of the department for limited resources, and to be able to manage and lead a diverse group of people (e.g., Sorcinelli, 1994). This diversity may be greater in geography than in many other disciplines because of the breadth of the discipline along the spectrum from science to social science and from qualitative to quantitative. The ideal chair would also need to have the ability to mentor and guide junior faculty as they commence their research and teaching careers. Being able to assist junior faculty as they attempt to find the balance between these often competing demands would be a valuable asset for the chair.

The attributes that contribute to creating a culture that values teaching would be similar at the faculty and the institution level, with minor additions. At the faculty and institutional level of many institutions there can be financial evidence of this teaching support with the creation of teaching fellowships, teaching awards and funding for scholarship of teaching and learning. There must be a perception of fairness among the stakeholders that the appropriate candidates are being successfully rewarded.

Teaching and Learning Centres have a valuable role to contribute in defining the culture of teaching at an institution. As well, they often reflect the current culture at the institution. Increasingly these centres must recognize the dual nature of their role. One part of the role truly is to provide remedial support to struggling teachers. Those

individuals that are struggling may reach the centre via direct referral or via self selection. The observation in this research project has been that the Centres only partly acknowledge this remedial role and prefer to assist individuals who have self-identified a need rather than providing a resource for all instructors who consistently score below some threshold on teaching assessments. Ideally, Teaching and Learning Centres need to be integrated into the administrative structure of the institution so that they are on similar footing with Research Services in order that these two activities are perceived as having equal footing at the institution. Both these areas should have a respective head that reports to the most senior level at the institution. Overall, geography chairs were very familiar with the staff, location and services provided by the Teaching and Learning Centre on their respective campuses. There was a great deal of breadth in the length of time these centres have existed. Some campuses have just gained a centre within the last three years, whereas other campuses have had the services of a centre for more than ten years.

Teaching culture is also influenced by discipline. As described in Section 3.5, geography is a discipline where teaching tends to be valued. Both Brown et al. (2002) and Cooke (1998) described the commitment and distinctiveness of geography teaching. The high value that geography as a discipline places upon teaching ensures that a supportive teaching culture can be developed within individual departments given faculty, institution and Teaching and Learning Centre support.

It is important to note the dashed line that separates culture from the evaluation environment. This dashed line illustrates two ideas. The first is that culture is dynamic and it is changing. The second, related idea, is that the evaluation environment will

influence the culture. These two ideas do not occur in isolation and there is feedback between the two, hence the dashed line.

i. Evaluation Environment

The demographics have been shown to influence the culture. These contribute to the idea that teaching evaluation does not occur in space, but rather occurs at a place. That place can be described as the evaluation environment. The evaluation environment simply represents the combination of the demographics and culture. It was the observation of this researcher that clarity about the demographics and culture would enable the development of an effective teaching evaluation system.

This chapter has described the synthesis of the knowledge gained from the interviews with geography department chairs, administrators, teaching and learning centre staff and a student leader. These results have been used to develop a revised conceptual framework for an effective teaching evaluation system. It has provided the reader an opportunity to more fully understand the system which can be applied at an institutional level to enhance the quality of existing teaching evaluation systems.

CHAPTER SEVEN:

CONCLUSION

7.1 Summary

The overall objective of this research was to identify good practices for teaching evaluation that could be applied within geography departments at Canadian universities. Ideally, the outcome of this work would be the identification of good practices for teaching evaluation that could be applied to other disciplines and within other countries. The overall objective of identifying good practices for teaching evaluation has been met with the creation of a model for a teaching evaluation system (Figure 6.1) which may have applicability beyond the discipline of geography and beyond the confines of Canada.

In order to meet this objective a series of research questions were identified. These were:

1. What is effective teaching in higher education?
2. What is effective teaching within the discipline of geography in higher education?
3. How and why is teaching evaluated in higher education?

4. What is the breadth of teaching evaluation practices currently used in geography departments within Canada?
5. How are the results of teaching evaluations used to enhance teaching quality within the discipline of geography in higher education?
6. How are the results of teaching evaluations used to reward teaching excellence within the discipline of geography in higher education?
7. What are 'good' teaching evaluation practices within the discipline of geography in higher education?

This research commenced with a thorough review of the literature on the characteristics of good teaching and effective teaching evaluation. This literature review provided the framework for the development of two conceptual frameworks (i.e., one for good teaching and one for effective teaching evaluation). These frameworks were tested within the discipline of geography through a national-level survey, interviews with geography department chairs and other stakeholders in the teaching evaluation system that were identified by the geography department chairs.

The reasons for evaluating teaching in higher education were explored both through the literature and within geography departments. The literature and the research participants identified that teaching evaluations are completed to ensure accountability, to provide instructors feedback, and to provide data for administrative decisions (e.g., promotion and tenure). Research participants also identified a purpose of evaluating teaching was to capture the student voice (i.e., perception) about teaching effectiveness. The literature identified two additional reasons for evaluating teaching that were not

described by research participants: data for curriculum improvement and design and data for research on teaching and learning.

Teaching is evaluated at three main stages in an instructor's career: at the time of hiring, for tenure and promotion, and at the end of each course. Nearly all departments with geography departments in Canada regularly evaluate teaching using course evaluations at the end of each course. There are no other formal methods of teaching evaluation that are regularly and widely used to evaluate teaching. Nearly all research participants emphasized the need to carefully consider a candidate's potential for teaching at the time of hire. Course evaluations are prone to concerns about credibility. Teaching evaluation practices within Canadian geography departments can be greatly enhanced by designing a broad system that considers all the stakeholders' needs and remains true to the underlying objective of enhancing student learning through teaching enhancements.

Feedback on teaching evaluations is used in a non-systematic way to enhance teaching quality. Although there appears to be an attempt to individualize the feedback and the mechanisms for the individual need of the instructor, it is quite unclear if there is much follow-up to ensure teaching quality actually improves. There was a glaring disconnect between the perceived role of teaching and learning centres between staff of the centres and chairs of Canadian geography departments.

The research has demonstrated that the conceptual framework informed by the literature on good teaching was further validated by the research data (see Figure 2.1). A total of 15 attributes of effective teaching were identified by chairs of geography. These attributes co-related well to those identified in the literature review.

The research has also demonstrated that the conceptual framework for effective teaching evaluation informed by the literature (see Figure 3.2) was found to be too simplistic. A revised framework based on the evidence gathered in this research is provided (see Figure 6.1). The revised framework describes the influence that the demographics and culture can have on the environment of evaluation. Place should be considered in designing an effective teaching evaluation system. There are seven parameters that contribute to the effective teaching evaluation system. These loosely match those discussed in the literature. The creation of this revised framework is an important contribution of this research.

7.2 Contributions

The major contribution of this work has been the development and testing of two conceptual frameworks. The first framework was for operationalizing effective teaching and the second was for a model for a teaching evaluation system. These models have applicability within Canadian geography departments and may have broader global implications. The thesis has allowed the researcher to explore and discuss the contribution of discipline to effective teaching and teaching evaluation. This work has necessitated the development of a research methodology that permits the exploration of a disciplinary definition of good teaching and good teaching evaluation. In addition this research has contributed to the field of geography teaching in higher education within Canada by promoting reflection about these topics.

7.3 Limitations

There are several limitations of the current project. This research project concentrated in one academic discipline (i.e., geography), within one country (i.e., Canada), used primarily one method of data collection (i.e., oral phone interviews) and studied one stakeholder predominantly (i.e., department chairs). Expanding the disciplines studied, the countries studied and the methodologies employed may provide a richer data set and more evidence to further test the conceptual frameworks. This would increase the credibility of the results.

The researcher has had the opportunity to spend the last 15+ years teaching in a research-intensive university setting. This background has provided the researcher the opportunity to deeply reflect on teaching and learning practices. These opportunities likely contributed in many positive ways to this research, but may also have led to higher than anticipated levels of researcher bias.

7.4 Future Directions

This research could be used to guide the creation of an operational tool to be used within institutions to guide the enhancement of an existing teaching evaluation system or to implement a teaching evaluation system. The research has clearly demonstrated that two key components of developing an effective teaching evaluation system are a thorough understanding of the demographics and culture; thus, the two key attributes to designing the system are geographical in nature. A case study research project could be designed around the operationalization of the teaching evaluation system.

As described in the section on limitations above, there were methodological limitations. Addressing those methodological limitations by studying more disciplines, in countries outside of Canada, using multiple methods of data collectionand expanding the study participants to include more of the stakeholders would enhance future research in this area. A future project could examine the applicability of the conceptual frameworks developed in this work to other disciplines and within other countries. Interviews were the primary method of data collection in this research. Exploring these conceptual frameworks using alternative methods of data collection (e.g., focus groups) may broaden our understanding of the parameters of both effective teaching and effective teaching evaluation. The primary stakeholder that was a research participant in this study was departmental chair. Engaging more of the stakeholders and examining the interconnections between their needs could be valuable future research. In the future it would be interesting to include more of the stakeholders of the teaching evaluation system in the research, particularly instructors (in non-administrative roles) and students. The credibility of a teaching evaluation system relies heavily on the perceptions of these two groups.

Although only a few teaching and learning centre staff participated in this research and a cursive scan of teaching and learning centre web sites was completed, there appeared in the research a disconnect between the users of the centre (e.g., chairs of geography departments) and the role of these centres perceived by the staff that work in the centres. Exploring this relationship more fully could be an interesting future project.

As Bob Dylan's famous song title states, "The Times, They are a Changin' " and this certainly appears to describe the situation within undergraduate education (Rae

Report, 2005; Boyer, 1996; Smith, 1991; Boyer, 1990). This research represents a snapshot in time. Since this change is occurring over a period of time a temporal project that examined changing perceptions and practices about effective teaching and teaching evaluation practices could be completed.

References

Abler, R. (1994). Toward a reconsideration of faculty roles and rewards in geography. *Journal of Geography in Higher Education, 18* (1), 9-18.

Andrews, J., Garrison, D. R., & Magnusson, K. (1996). The teaching and learning transaction in higher education: A study of excellent professors and their students. *Teaching in Higher Education, 1* (1), 81-103.

Angelo, T. A. (1996). Relating exemplary teaching to student learning. *New Directions for Teaching and Learning, 65*, 57-65.

Arreola, R.A. (1986). Evaluating the dimensions of teaching. *Instructional Evaluation, 8* (2), 4-12.

Arreola, R.A. (1989). Defining and evaluating the elements of teaching. In W. Cashin (Ed.) *Proceedings of academic chairpersons: Evaluating faculty, students and programs* (pp. 3-12). Manhattan, KS: Kansas State University.

Arreola, R.A. (2000). *Developing a Comprehensive Faculty Evaluation System: A Handbook for College Faculty and Administrators on Designing and Operating a Comprehensive Faculty Evaluation System.* Bolton, MA: Anker Publishing.

Association of Universities and Colleges of Canada. (2007). *Our universities.* Retrieved November 26, 2007, from http://www.aucc.ca/can_uni/our_universities/index_e.html

Babbie, E. (2001). The *Practice of Social Research.* Belmont, CA: Wadsworth/Thomson Learning.

Basow, S. A. (1994, June 21). *Student ratings of professors are not gender blind.* Message posted to Forum for Teaching and Learning in High electronic mailing list, archived at https://listserv.unb.ca/cgi-bin/wa?A2=ind9406&L=stlhe-l&O=D&P=3335.

Basow, S.A. & Distenfeld, M.S. (1985). Teacher expressiveness: More important for male teachers than female teachers? *Journal of Educational Psychology, 77* (1), 45-52.

Beeby, C.E. (1977). The Meaning of Evaluation. Current Issues in Education: No. 4 Evaluation Wellington: Department of Education, p. 68-78.

Boyer, E. L. (1990). *Scholarship reconsidered: Priorities of the professoriate.* Lawrenceville, NJ: Princeton, Carnegie Foundation for the Advancement of Teaching.

The Boyer Commission on Educating Undergraduates in the Research University. (1998). *Reinventing undergraduate education: A blueprint for America's research universities.* Stony Brook, NY: Author.

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

Braskamp, L. A., & Ory, J. (1994). *Assessing faculty work: Enhancing individual and institutional performance.* San Francisco: Jossey-Bass.

Brown, S. (1965). *Social psychology.* New York: New York Free Press.

Brown, S., Bucklow, C., & Clark, P. (2002). Professionalising teaching: Enhancing the status of teaching, improving the experience of learning and supporting innovation in higher education. *Journal of Geography in Higher Education, 26* (2), 159-168.

Burkill, S. (2002). Recognizing and rewarding excellent teachers: Toward a strategy for geography departments. *Journal of Geography in Higher Education, 26* (3), 253-262.

Carini, M., Kuh, G., & Klein, S. (2006). Student engagement and student learning: Testing the linkages. *Research in Higher Education, 47* (1), 1-32.

Carroll, L. (2000). *Alice's adventures in Wonderland.* Cambridge, UK: Bell and Howell. (Original work published 1866). Retrieved January 21, 2008, from Literature Online (LION) database.

Cashin, W. E. (1988). Idea Paper No. 21: Student ratings of teaching: A summary of the research. *Idea Paper, 20.*

Cashin, W. E. (1989). *Defining and evaluating college teaching.* Idea Paper No. 21. Manhattan, KA: Kansas State University.

Cashin, W. E. (1995). Student ratings of teaching: The research revisited. Idea Paper No. 32. Manhattan, KA: Kansas State University.

Cashin, W. E. (1996). Developing an effective faculty evaluation system. Idea Paper No. 33. Manhattan, KA: Kansas State University

Canadian Association of University Teachers. (2004). *Canadian Association of University Teachers Almanac 2004.* Retrieved July 4, 2004, from http://www.caut.ca/english /publications/cautalmanac/default.asp.

Canadian Association of University Teachers. (2007). CAUT almanac of post-secondary education in Canada 2007. Ottawa, ON: Author.

Centra, J. A. (1977). *How universities evaluate faculty performance: A survey of department heads* (GREB No.75-5bR). Princeton, NJ: Educational Testing Centre.

Centra, J. A. (1979). *Determining faculty effectiveness: Assessing teaching, research, and service for personal decisions and improvement*. San Francisco: Jossey-Bass.

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

Centra, J.A. (1994). Current issues in evaluating and improving college teaching. Paper presented at American Educational Research Association annual meeting, Atlanta, GA.

Chalkley, B. (1998). Geography and quality of higher education. *Geography, 83* (358), 53-62.

Chalkley, B., Fournier, E. J., & Hill, A. D. (2000). Geography teaching in higher education: quality, assessment, and accountability. *Journal of Geography in Higher Education, 24* (2), 238-245.

Chickering, A. & Gamson, Z. (1989). Seven principles for good practice in undergraduate education. *AAHE Bulletin, 3*, 3-7.

Cohen, P. (1980). Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. *Research in Higher Education*, 23, 321-341.

Cohen, P. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research, 51* (3), 281-309.

Cooke, R. U. (1998). Enhancing teacher quality. *Journal of Geography in Higher Education, 22* (3), 283-284.

Cox, M. D. (2000). Students can help us improve our teaching. *The Teaching Professor, 14* (2), 1-3.

Crebbin, W. (1997). Teaching for lifelong learning. In R. Ballantyne, J. Bain, & J. Packer (Eds.), *Reflecting on university teaching: Academics' stories* (pp. 139-150). Canberra: Australian Government Publishing Service.

Derry, J. O. (1979). Can students' ratings of instruction serve rival purposes? *Journal of Higher Education, 50* (1), 79-88

Diamond, R., & Adam, B. (1993). *Recognizing faculty work: Reward systems for the year 2000.* New directions for higher education no. 81. San Francisco, Jossey-Bass.

Diamond, R., & Adam, B. (1995). *The disciplines speak: Rewarding the scholar, professional, and creative work of faculty.* Washington, DC: American Association for Higher Learning.

Druckers, A. J., & Remmers, H. H. (1951). Do alumni and students differ in their attitude toward instructors? *Journal of Educational Psychology, 42*, 129-143.

Edgerton, R. (1993). The re-examination of faculty priorities. *Change, 25* (4), 10-25.

Edgerton, R., Hutchings, P., & Quinlan, K. (1991). *The teaching portfolio: Capturing the scholarship in teaching.* Washington, DC: American Association for Higher Education.

Elton, L. (1998). Dimensions of excellence in university teaching. International Journal for *Academic Development, 3* (1), 3-11.

Falcigno, K. (2005). *The Canadian Association of Geographers Directory 2005.* Montreal: Canadian Association of Geographers.

Falcigno, K. (2006). *The Canadian Association of Geographers Directory 2006.* Montreal: Canadian Association of Geographers.

Falcigno, K. (2007). *The Canadian Association of Geographers Directory 2007.* Montreal: Canadian Association of Geographers.

Farrington, F. (2000). Benchmark standards for geography: A personal review. *Journal of Geography in Higher Education, 24* (3), 413-415.

Feldman, K. A. (1976). The superior college teacher from the students' view. *Research in Higher Education, 5* (3), 243-288.

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

Feldman, K. A. (1979). The significance of circumstances for college students' ratings of their teachers and college. *Research in Higher Education, 10* (2), 149-172.

Feldman, K. A. (1983). Seniority and experience of college teachers as related to evaluation they receive from their students. *Research in Higher Education, 18* (1), 3-124.

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

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

Feldman, K. A. (1988). Effective college teaching from students' and faculty's view: Matched or mismatched priorities? *Research in Higher Education, 28* (4), 291-344.

Feldman, K. A. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education, 30* (6), 583-645.

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

Feldman, K., & Paulsen, M. (1999). Faculty motivation: The role of a supportive teaching culture. *New Directions for Teaching and Learning, 78*, 71-79.

Geography Benchmarking Group. (2000). The Quality Assurance Agency's benchmark statement for geography. *Journal of Geography in Higher Education, 24* (3), 399-412.

Gibbs, G. (1999). Improving teaching, learning, and assessment. *Journal of Geography in Higher Education, 23* (2), 147-155.

Gold, J. R., Jenkins, A., Lee, R., Monk, J., Riley, J., Sheperd, I, et al. (1991). *Teaching geography in higher education: A manual of good practice.* Cambridge, MA: Basil Blackwell.

Hannerz, U (1993). *Cultural Complexity – Studies in the Social Organization of Meaning.* Columbia, NY: Columbia University Press.

Hattie, J. (1992). *Self concept.* Hillsdale, NJ: Lawrence Erlbaum.

Healey, M. (2000). Developing the Scholarship of Teaching in Higher Education: a discipline-based approach. *Higher Education and Development,* 19 (2), 169-189.

Higher Education Quality Control Council of Ontario. (2005). *The Higher Education Quality Council of Ontario Act.* Toronto: Author.

Higher Education Quality Control Council of Ontario. (2007). *Higher Education Quality Council of Ontario Annual Report 2006/2007.* Toronto: Author.

Johnson, A. D. (2003, November 17). Measuring excellence: An insider's guide [Special issue: Universities 2003]. *Maclean's.* pp.26-119.

Knapper, C. K., & Rogers, P. (1994). Increasing the emphasis on teaching in Ontario universities. Toronto, ON: Ontario Council on University Affairs Task Force on Resource Allocation.

Kulik, J. A., & McKeachie, W. J. (1975). The evaluation of teachers in higher education. In F. N. Kerlinger (Ed.), *Review of research in higher education* (Vol. 3, pp. 210-240). Itasca, IL: Peacock.

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

Lewin, K. (1951). *Field Theory in Social Science.* Harper and Row, New York.

Marantz, H., & Warren, A. (1998). A conservative view of geographical education. *Journal of Geography in Higher Education, 22* (1), 49-53.

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

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

Marsh, H. W. (1990). A multidimensional, hierarchical model of self concept: Theoretical and empirical justification. *Educational Psychology Review, 2* (2), 77-172.

Marsh, H. W., & Craven, R. G. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement and adjustment* (pp. 131-198). San Diego, CA: Academic Press.

Marsh, H. W., & Overall, J. U. (1979). Long-term stability of students' evaluations: A note on Feldman's "Consistency and variability among college students in rating their teachers and courses". *Research in Higher Education, 10* (2), 139-147.

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

Massy, W., Wilger, A., & Colbeck, C. (1994). Overcoming "hollowed" collegiality. *Change,26* (4), 11-20.

McKeachie, W. J. (1958). *The appraisal of teaching in large universities.* Ann Arbor, MI: University of Michigan Report.

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

Menges, R. J. (1973). The new reporters: Students rate instruction. In C. R. Pace (Ed.), *Evaluating learning and teaching: New directions for higher education* (pp. 59-75). San Francisco: Jossey-Bass.

Moore, G. (1999). *Agriculture 735 Online Site.* Retrieved June 27, 2004, from http://www.cals.ncsu.edu/agexed/aee735/ppt1/.

Murray, H. G. (1973). *A guide to teaching evaluation.* Toronto: Ontario Confederation of University Faculty Associations.

Murray, H. G. (1980). A comprehensive plan for the evaluation of teaching at the University of Queensland. *Reflections, 4*, 3-11.

Murray, H. G., Newby, W., Crealock, C., Bowden, B., Gailey, T., Oswin, J., et al. (1982). *Evaluation of teaching at the University of Western Ontario.* London, ON: University of Western Ontario.

Murray, H. G., & Renaud, R. D. (1995). Disciplinary differences in classroom teaching behaviors. *New Directions for Teaching and Learning, 64*, 31-39.

National Survey of Student Engagement. (2007). *Our origins and potential.* Retrieved January 19, 2007, from http://www.nsse.iub.edu/html/origins.cfm.

Nelson, R., & Semple, S. (2000). Cross campus cooperation for geographers: A summary of questionnaire results. *Canadian Association of Geographers Newsletter, 7* (5), 10-12.

Newcombe, T.M., Turner, R.H., & Converse, P.E. (1965). *Social psychology: The study of human interaction.* New York: Holt, Rinehart, and Winston.

O'Connell, S., Ortiz, J., & Morrison, J. (2003). Connecting with the river. *Geotimes, 48* (9), 14-17.

Overall, J. U., & Marsh, H. W. (1980). Students' evaluations of instruction: A longitudinal study of their stability. *Journal of Educational Psychology, 72* (3), 321-325.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research.* San Francisco: Jossey-Bass.

Pinsky, L. E., & Irby, D. M. (1997). If at first you don't succeed: Using failure to improve teaching. *Academic Medicine, 72* (11), 973–976.

Prensky, M. (2001). Do they Really *Think* Differently? *On the Horizon* 9 (6), 1-9.

Rae, B. (2005). *Ontario: A leader in learning*. Toronto: Publications Ontario.

Ramsden, P. (1979). Student learning and perceptions of the academic environment. *Higher Education, 8* (4), 411-427.

Ramsden, P., & Entwistle, N. (1981). Effects of academic departments on students' approaches to studying. *British Journal of Educational Psychology, 51* (3), 368-383.

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The course experience questionnaire. *Studies in Higher Education, 16* (2), 129-150.

Ramsden, P. (1992). *Learning to teach in higher education.* London: Routledge.

Rice, R. E. (1986). The academic profession in transition: Toward a new social fiction. *Teaching Sociology, 14*, 12-23.

Robinson, T. A. (2001, August). *Government of Canada caucus on post secondary education and research.* Paper presented at a meeting of the Confederation of Alberta Faculty Associations, Edmonton, AB.

Roche, L. A. and Marsh, H. W. (2000). Multiple dimensions of university teacher self-concept. *Instructional Science, 28* (5), 439-468.

Schank, R. C. (2002). Are we going to get smarter? New York: Vintage Books.

Seldin, P. (1980). *Successful faculty evaluation programs*. Crugers, NY: Coventry Press.

Seldin, P. (1984). *Changing practices in faculty evaluation: A critical assessment and recommendations for improvement*. San Francisco: Jossey-Bass.

Seldin, P. (1991). *The teaching portfolio: A practical guide to improved performance and promotion/tenure decisions*. Boston: Anker.

Shulman, L. (1993, November/December). Teaching as community property: Putting an end to pedagogical solitude. *Change, 25* (6), 6-7.

Smith, B. (2000). When good teaching becomes outstanding: The pursuit of excellence. *Educational Developments, 1* (20), 6-7.

Smith, D. M. (1988). On academic performance. *Area, 20*, 3-13.

Smith, S. L. (1991). *Report: Commission of inquiry on Canadian university education.* Ottawa: Association of Universities and Colleges of Canada.

Snyder, V. R. (1971). *The hidden curriculum.* New York: Knopf.

Sorcinelli, M.D. (1994). Effective Approaches to Faculty Development. *Journal of Counselling and Development* 72 (5), 474-479.

Statistics Canada. (1998). *Population 15 years and over by sex (3) and total income groups (22)in constant (1995) dollars, for Canada, provinces, territories and census metropolitan areas, 1990 and 1995 (20% Sample Data).* Retrieved June 26, 2004, from http://www.statcan.ca:8096/bsolc/english/bsolc?catno=93F0029X1996001.

Statistics Canada. (2005). Projected population by age group and sex according to a medium growth scenario for 2006, 2011, 2016, 2021, 2026 and 2031, at July 1. Retrieved November 21, 2007, from http://www40.statcan.ca/101/cst01/demo23a.htm?sdi=ages.

Sullivan, A. M., & Skanes, G. R. (1974). Validity of student evaluation of teaching and the characteristics of successful instructors. *Journal of Educational Psychology, 66* (4), 584-590.

Teaching. (2004). In *Merriam-Webster Dictionary Online.* Retrieved June 28, 2004, from http://www.m-w.com/cgi-bin/dictionary.

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

Tricart, J. (1969). *The teaching of geography at university level.* London: George G. Harrap.

Walls, R. T., Nardi, A. H., von Minden, A. M., & Hoffman, N. (2002). The Characteristics of Effective and Ineffective Teachers. *Teacher Education Quarterly, 29* (1), 39-48.

Watkins, D., & Akande, A. (1992). Student evaluations of teaching effectiveness: A Nigerian investigation. *Higher Education, 24* (4), 453-463.

Wilson, R. (1986). Improving faculty teaching: Effective use of student evaluations and consultants. *Journal of Higher Education, 57* (2), 196-211.

Wolf, R. (1987). The nature of educational evaluation. *International Journal of Educational Research, 11* (1), 7-19.

Woolf, H. (2004). Assessment criteria: Reflections on current practices. *Assessment and Evaluation in Higher Education, 29* (4), 479-493.

Wright, A. (1999, January/February). Improving teaching by design. *Focus on University Teaching and Learning, 8* (2), 1-7.

Wright, W. A., & O'Neil, C. (1992, January). Student ratings of instruction: Principles for practice. *Focus on University Teaching and Learning, 4*, 1-4.

Young, S., & Shaw, D. (1999). Profiles of effective college and university teachers. *Journal of Higher Education, 70* (6), 670-686.

APPENDIX ONE:

Ethics Certificate National Survey

APPENDIX TWO:

National Survey Instrument

Yes
No

If no, please continue to question 5. If yes, please continue to question 4.

4. Please complete the following table with graduate enrolment data from your department as of September, 2004.

Name of Program	**Full-time or Part-time**	**Enrolment**
M.A.	Full-time	
M.A.	Part-time	
M.Sc.	Full-time	
M.Sc.	Part-time	
PhD.	Full-time	
PhD.	Part-time	
Other: please specify		

5. What was the total number of undergraduate student equivalents in your department during the 2004-2005 academic year (i.e. September 2004 to April 2005): _____?

6. How many students received an undergraduate degree from a program offered by your department during 2004-2005 (please specify the name of the program, whether the program was an honours program and the number of students):

Name of Program	**Honours (Y/N)**	**# of Students**

Part 2: Undergraduate Course Information

This second part of the survey is asking a series of questions to determine information about the courses offered by your department

7. How many undergraduate courses were taught in 2004-2005 in your department:

_____

8. Of the courses identified in #7 above, how many were full year courses?

_____

9. What percentage of your full year courses is taught by tenured/tenure track faculty members? _____

10. What percentage of your half year courses is taught by tenured/tenure track faculty members? _____

11. Does your institution offer on-line undergraduate geography courses for credit? Please circle the correct response.

YES NO

If no, please continue to question 14. If yes, please continue to question 12.

12. How many on-line undergraduate geography courses do you offer annually for credit?

_____

13. What percentage of your online undergraduate geography courses is taught by tenured/tenure track faculty members? _____

Part 4: Undergraduate Program Information

Part four of the survey is asking a series of questions about your undergraduate programs.

14. What undergraduate degrees do you offer in your department (please list if they are honours (four year or five year) or if they are general(pass) degrees (three year) and if they are BA or BSc degrees)?

Name of program	Length of Degree (e.g. three years)	Honours or General	B.A. or B.Sc. or ?	Enrolment 2004-2005

15. In what year do students register in your programs (level one or level two)?

_____

Part 5: Instructional Staff Information

The fifth section of the survey asks a series of questions about the instructional staff that are involved in teaching your students.

16. Please complete the table below to describe your faculty complement as of September 1, 2004.

Rank	Total Number	Female	Male
Professor			
Associate Professor			
Assistant Professor			
Lecturer			

Sessional			
Other: please specify			

17. Are your faculty members of a legally recognized union? Please circle the correct response.

YES NO

If no, please continue to question 19. If yes, please continue to question 18.

18. What is the union's name: _____ Please continue to question 21

19. Are your faculty represented by an association? Please circle the correct response.

YES NO

20. What is the association's name: _____

21. What percentage of the full-time graduate students in your department have TA positions?

_____

22. Are your graduate students represented by a union in their TA postions? Please circle the correct response.

YES NO If no please skip to question 24.

23. What is the name of your graduate students TA union?

_____

24. Please complete the table below to describe your teaching support positions as of September 1, 2004.

Position	**Total Number**	**Female**	**Male**
Graduate Student Teaching Assistants (TAs)			
Undergraduate			

Student TAs			
Other TAs			
Markers			
Full time Instructional Assistant Positions			
Other: Please specify			

Part 6: Teaching Evaluation Process

The following section asks a series of questions to determine information about the process(es) used in your department/faculty/institution to evaluate teaching.

25. Why do you evaluate teaching (please circle all that apply)?

a. to provide diagnostic feedback to faculty about their teaching

b. as a measure to be used for administrative purposes to assist in guiding decisions about promotion, tenure and salary

c. to provide information to prospective students to assist them in making course selections

d. to asses the quality of individual courses to be used for course and curriculum improvement and design

e. to provide data for research on teaching and learning

f. other (please specify): _____

26. Please complete the following table identifying the measurement tool(s) that you use to evaluate teaching for each of the reasons identified in question 25 above.

Reason to Evaluate		**Diagnostic Feedback**	**Administrative Purposes**	**Student Information**	**Course & Curriculum Design**	**Research on Teaching & Learning**	**Other: please specify**
Measurement Tool	Student Evaluation – standard form						
	Student Evaluation – feedback other than that solicited on form						
	Instructor Self Reflective Evaluation						
	Peer/Colleague Evaluation						
	Supervisor Evaluation						
	Evidence of Methods to Contribute to Effective Teaching						
	Teaching Portfolio						
	Other: please describe						

27. For each of the tools used to measure teaching effectiveness complete the following table to comment on the frequency of use.

Frequency of Use		**Completed for every course**	**Completed annually**	**Completed as needed by instructor**	**Completed as needed by administration**	**Completed for tenure & promotion decisions**	**Other: please specify**
Measurement Tool	Student Evaluation – standard form						
	Student Evaluation – feedback other than that solicited on form						
	Instructor Self Reflective Evaluation						
	Peer/Colleague Evaluation						
	Supervisor Evaluation						
	Evidence of Methods to Contribute to Effective Teaching						
	Teaching Portfolio						
	Other: please describe						

28. Nearly every institution within Canada, US, UK and Australia uses a student evaluation of teaching effectiveness form on which students evaluate the quality of their instruction in each of their courses. If you do not use one of these forms please skip to question 36. If you do use one of these forms please complete the following series of questions about the form.

a. Do you use a standard student evaluation of teaching effectiveness form in your (please check all that apply):

i. department
ii. faculty
iii. institution

b. Who was responsible for the design of the student evaluation of teaching effectiveness form?

_____

c. Is the form based upon an existing published student evaluation of teaching effectiveness form? Please circle the correct answer

YES NO; If no please continue with d

If yes, what form?

_____

d. How often is the student evaluation of teaching effectiveness form updated/changed?

_____

e. What is the process for making changes to the student evaluation of teaching effectiveness form?

_____

_____

_____

_____

f. Who has access to the results from the student evaluation of teaching effectiveness forms (please circle all that apply)?

i. chair
ii. dean
iii. instructor
iv. students
v. other: please specify

g. Is the student evaluation of teaching effectiveness form completed online? Please check the correct answer.

YES NO; please continue with h

If yes, comment on the success of the online form.

_____

_____

_____

_____

h. The numerical data on the student evaluation of teaching effectiveness forms is usually aggregated before the instructor sees the results. How is the aggregated data presented? Please circle all that apply.

i. means
ii. medians
iii. modes
iv. standard deviation
v. other: please specify

i. What is the average response rate obtained on the student evaluation of teaching effectiveness forms?

29. Many departments also evaluate their teaching assistants. Does your department evaluate teaching assistants? Please circle the correct answer.

YES, please continue with this question
NO, please continue with question 30

a. Do you use a standard teaching assistant evaluation form in your (please check all that apply):

i. department
ii. faculty
iii. institution

b. Who was responsible for the design of the teaching assistant evaluation form?

_____

c. Is the form based upon an existing published teaching assistant evaluation form? Please circle the correct answer.

YES NO; please continue with d

If yes, what form?

_____

d. How often is the teaching assistant evaluation form updated/changed?

_____

e. What is the process for making changes to the teaching assistant evaluation form?

_____

_____

_____

_____

f. Who has access to the results from the teaching assistant evaluation forms (please circle all that apply)?

i. chair
ii. dean
iii. instructor of the course
iv. students
v. teaching assistant
v. other: please specify

g. Is the form completed online? Please circle the correct answer.

YES NO; please continue with h

If yes, comment on the success of the online form.

_____

_____

_____

_____

h. The numerical data on the forms is usually aggregated before the teaching assistant sees the results. How is the aggregated data presented?

i. mean
ii. median
iii. mode
iv. standard deviation
v. other: please specify

i. What is the average response rate obtained on the teaching assistant evaluation forms?

Part 7: Perspectives

This section asks for feedback based on your perspectives about the teaching evaluation scheme used at your institution.

30. What do you think are the strengths of the teaching evaluation scheme used in your unit?

31. What do you think are the weaknesses of the teaching evaluation scheme used in your unit?

32. How do you think the teaching evaluation scheme used in your unit could be enhanced?

Part 8: Request for Information

In this section you are being asked to provide the written documentation that exists for your unit about teaching evaluations. I had originally thought that I could obtain this information from institution's web sites, but, I have not had success accessing the materials off of web sites.

Please provide copies of any documentation that you may have that describes teaching evaluation practices at your institution and in your unit. This may include, but is not limited to,

union by-laws on teaching evaluation
faculty handbook about teaching evaluation
senate or university policy on teaching evaluation
unit policy on teaching evaluation
copy of the student evaluation form on instruction
copy of the student evaluation form on teaching assistants.

THANK YOU for taking the time to complete this survey. The results of this survey and the follow-up will be presented at the Annual Meeting of the Canadian Association of Geographers (CAG) in 2007. Preliminary results will be presented at the CAG meeting in 2006. Each department that completes a survey will receive a summary of the results once the thesis is completed.

Please provide any additional information that you feel may be relevant.

APPENDIX THREE:

Ethics Certificate Oral Interviews

APPENDIX FOUR:

Oral Interview Survey Instruments

Oral Interviews with Chairs

1. How is undergraduate teaching evaluated?
    a. at your institution
    b. within your department

include a list of items that are potentially used as probative: Do you use any of the following? How? When?

- *self evaluation*
- *student evaluation*
    - *course evaluations*
    - *student recorders/observers*
    - *faux student in the class*
    - *student videotape*
    - *interviews with students as focus groups*
    - *using trained student consultants*
- *peer/colleague evaluation*
- *supervisor evaluation (e.g. Chair, Director, Dean)*
- *evidence of teaching methods that have demonstrated the ability to contribute to effective student learning*
- *teaching portolio/dossiers*

2. What is the purpose of teaching evaluation?
    a. at your institution
    b. within your department

3. What is done with the results of teaching evaluations?
    a. as a process
    b. to reward teaching excellence
    c. to enhance teaching quality

4. What part(s) of the teaching evaluation process do you think works best?

5. What part(s) of the teaching evaluation process do you think works least well?

6. How is teaching rewarded?
    a. at your institution
    b. within your department

7. How is teaching enhanced?
    a. at your institution
    b. within your department

8. How would you define good teaching?

9. How would you define good teaching?

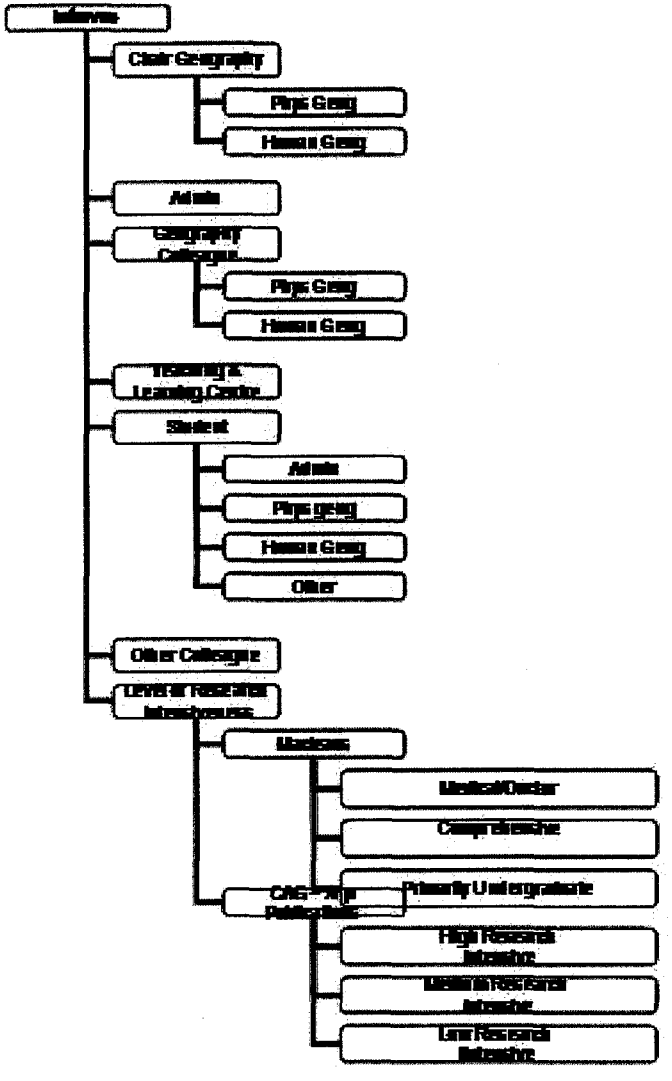
10. Do the instruments used to measure teaching effectiveness at your _____ measure your definition of good teaching? Explain?

11. Are there any people at your institution that you would suggest that I speak to in order to better understand the teaching evaluation process?

12. Comments? Questions?

APPENDIX FIVE:

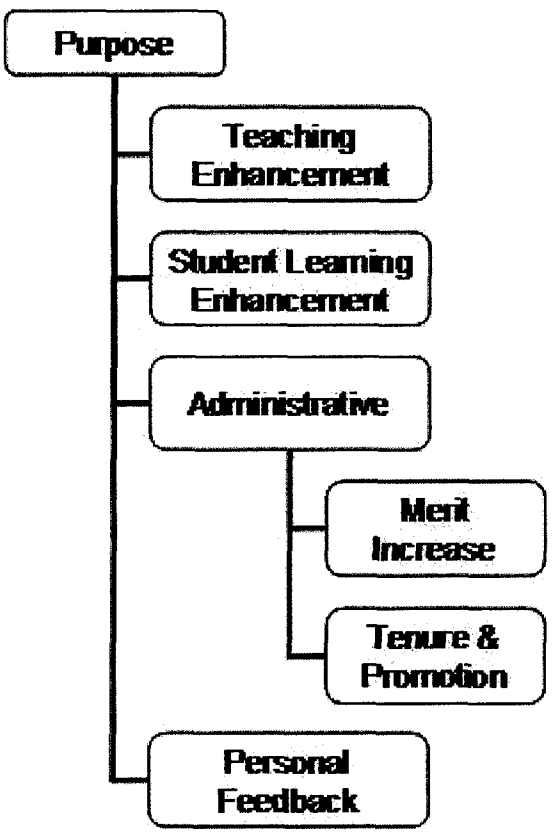
NVIVO Codes from oral interviews



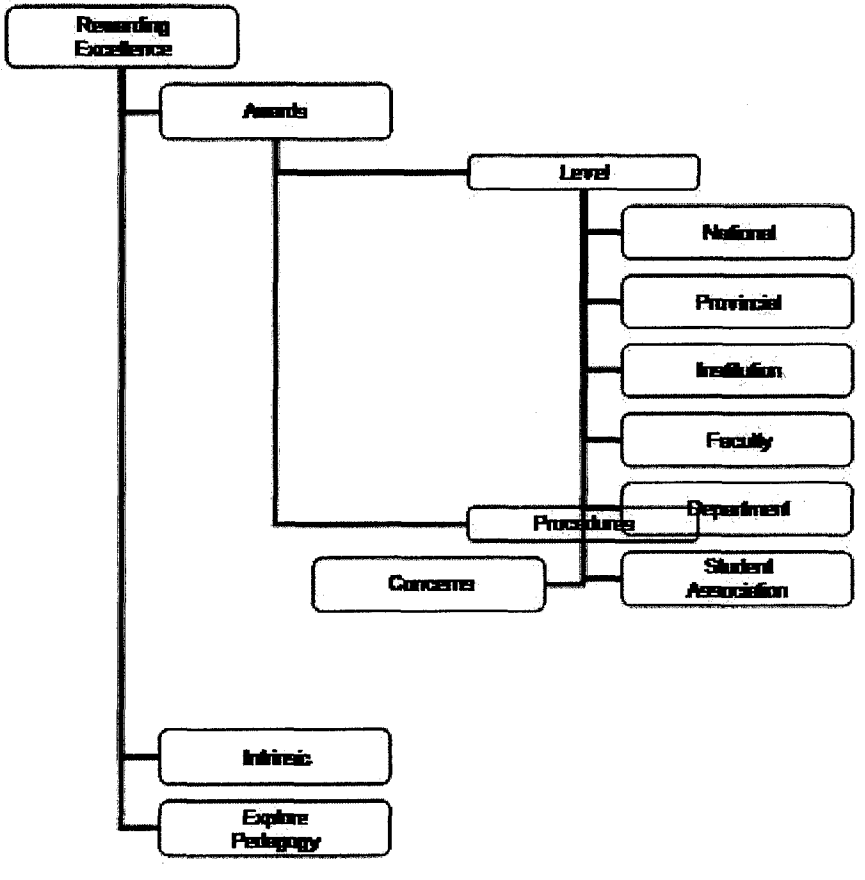
Node Structure: Research Interviews

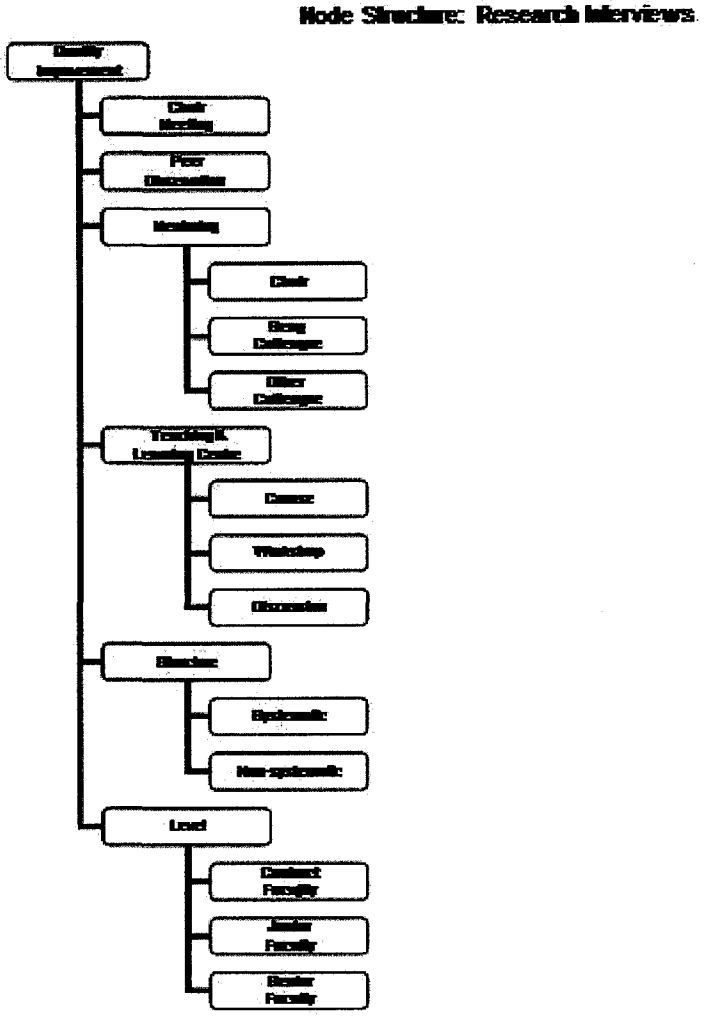
Node Structure: Research Interviews

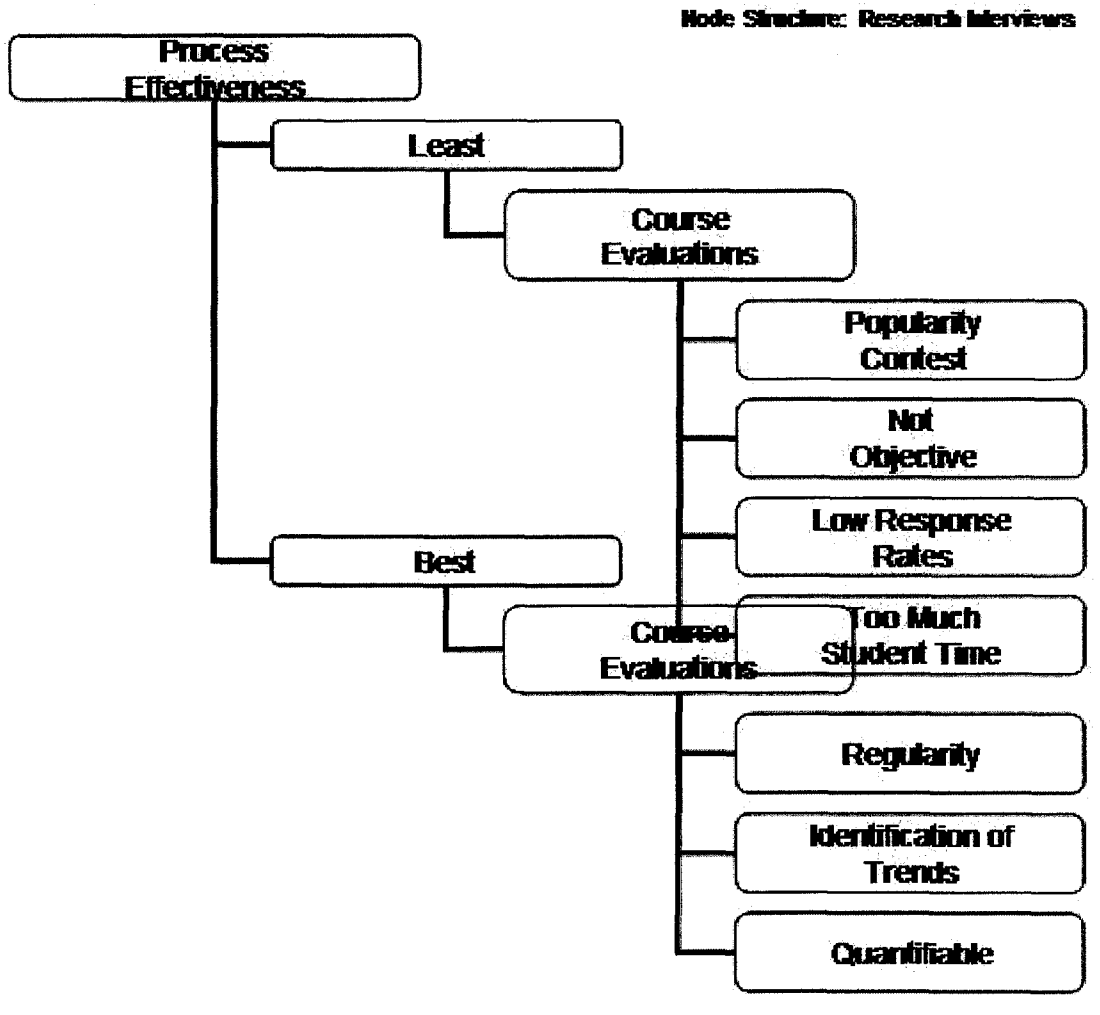
Node Structure: Research Interviews

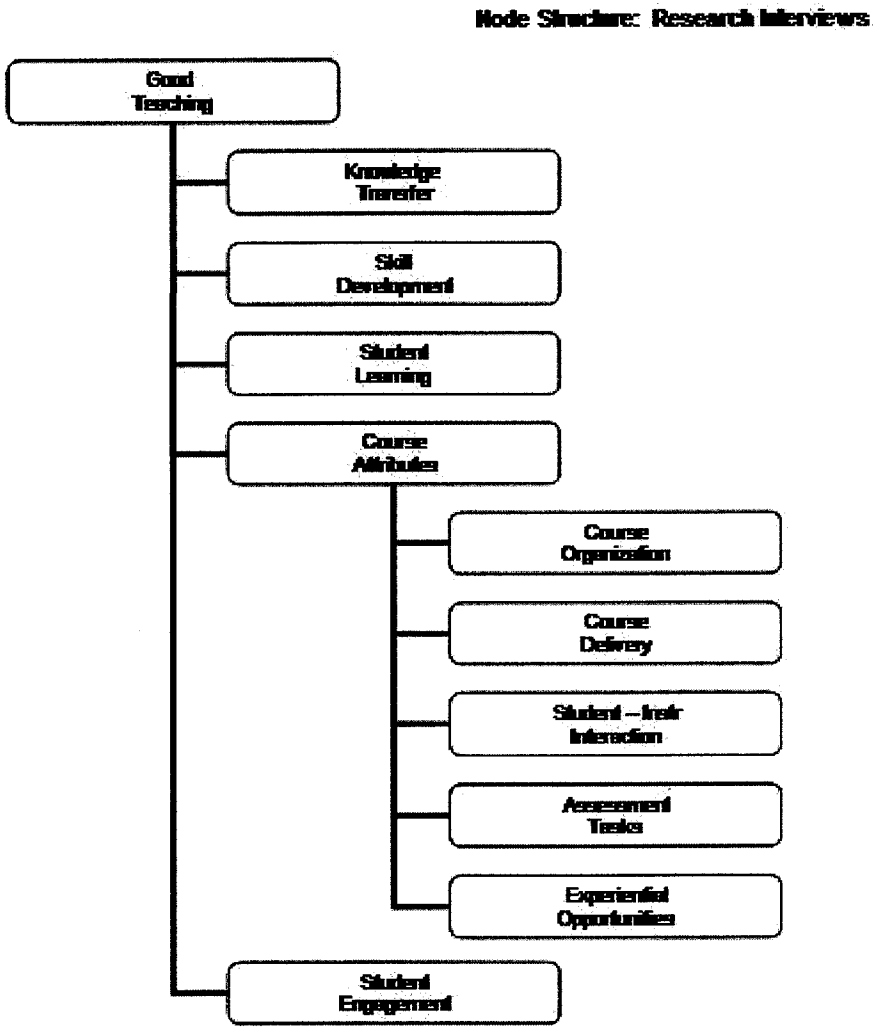


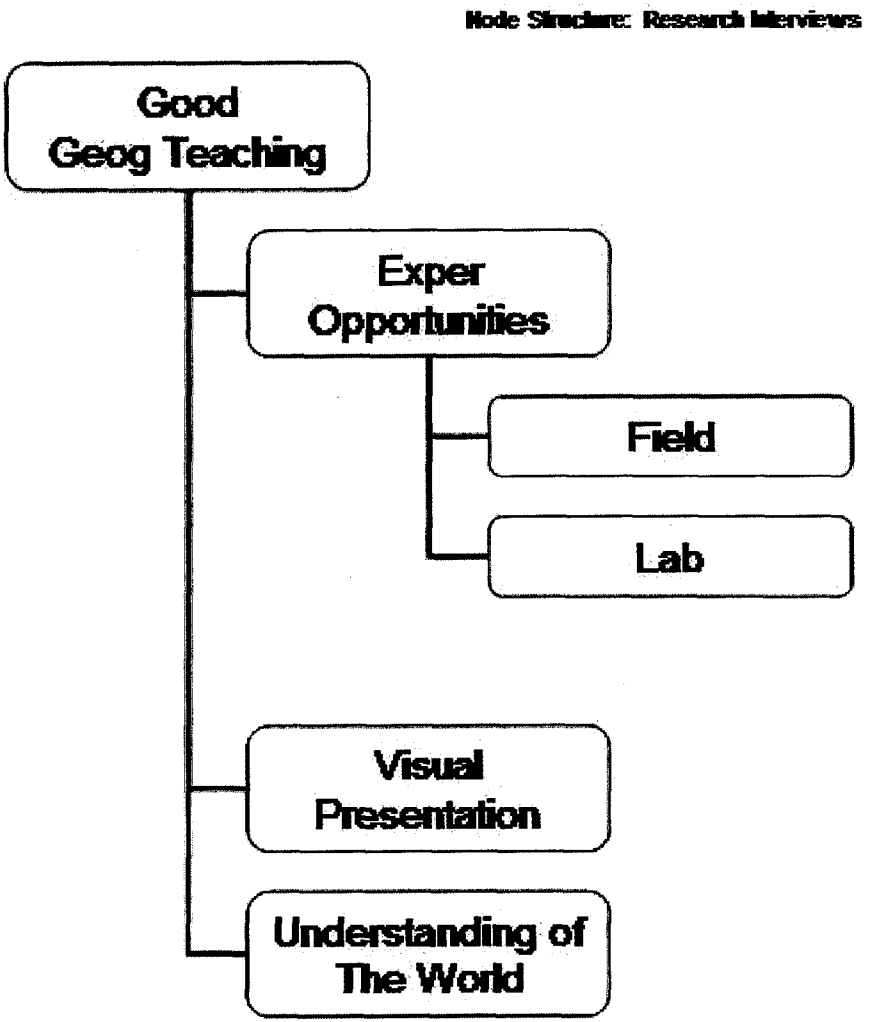
Node Structure: Research Interviews



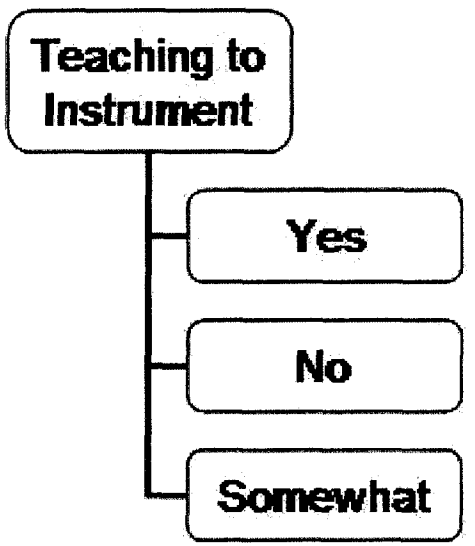

Node Structure: Research Interviews

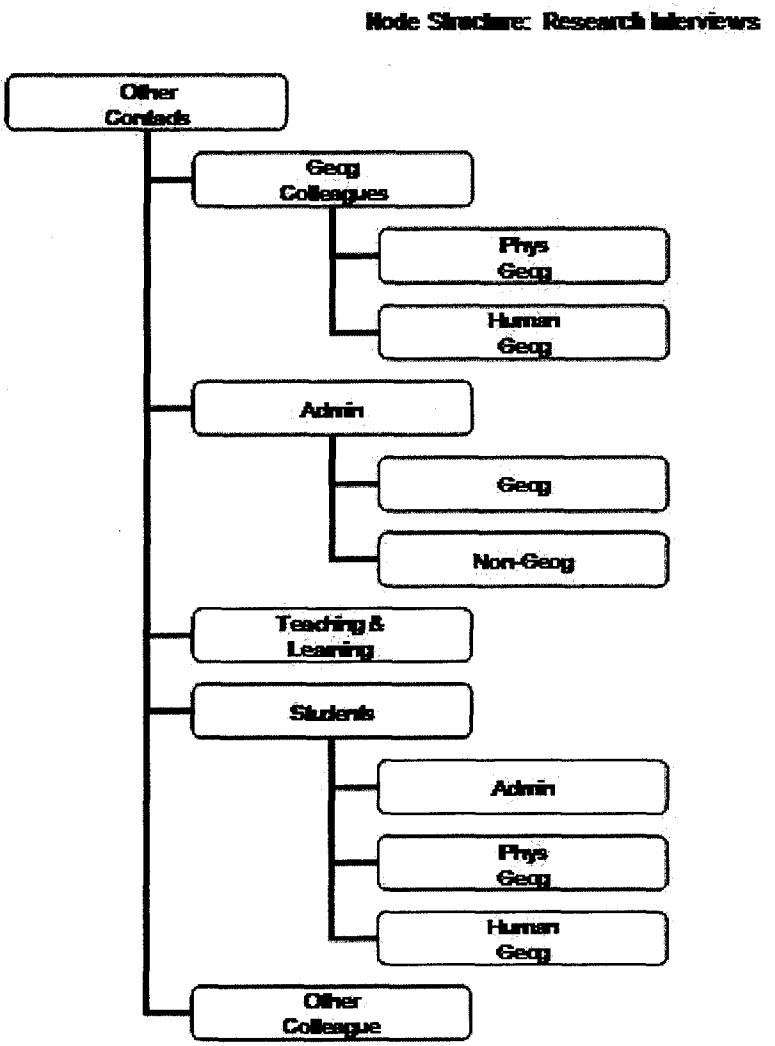

Node Structure: Research Interviews
Process Effectiveness
Least
Course Evaluations
Popularity Contest
Not Objective
Low Response Rates
Too Much Student Time
Best
Course Evaluations
Regularity
Identification of Trends
Quantifiable


Node Structure: Research Interviews
Good Teaching
Knowledge Transfer
Skill Development
Student Learning
Course Attributes
Course Organization
Course Delivery
Student – Instr Interaction
Assessment Tasks
Experiential Opportunities
Student Engagement


Node Structure: Research Interviews
Good Geog Teaching
Exper Opportunities
Field
Lab
Visual Presentation
Understanding of The World

Node Structure: Research Interviews




Node Structure: Research Interviews
Other Contacts
Geog Colleagues
Phys Geog
Human Geog
Admin
Geog
Non-Geog
Teaching & Learning
Students
Admin
Phys Geog
Human Geog
Other Colleague