THE DERIVATION AND APPLICATION OF STANDARDS IN DISTANCE EDUCATION PROGRAM EVALUATION

NADINE FLANNIGAN-WHEELER

THE DERIVATION AND APPLICATION OF STANDARDS IN DISTANCE EDUCATION PROGRAM EVALUATION

by

Nadine Flannigan-Wheeler, B.Sc. (Hon.)

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Master of Education

Faculty of Education
Memorial University of Newfoundland
August 1996

ABSTRACT

The purpose of this study was to explore the program evaluation methods employed at an autonomous distance education institution, focusing in particular on the derivation and application of standards within the evaluation process. This study also attempted to examine the relationship between the data gathered and theory and models in the evaluation literature. The study was initiated to provide baseline data on the application and derivation of standards as this subject is not addressed extensively in the literature.

Using an interpretive case study approach this study was implemented in the summer of 1995, and employed semi-structured to unstructured interviews, questionnaires and document analysis to elicit information. All interviews permitted open responses, and were tape-recorded with the permission of respondents.

The data were initially analyzed using Kripendorff"s (1980) semantic content analysis resulting in a narrative report. This was followed by the development of categories and themes as suggested by Merriam (1988) in order to

interpret and theorize about the data with reference to the literature.

Results of the study indicated that no one approach to the setting and use of standards and criteria is employed. Input from stakeholders, the development of performance indicators by departments, university policy, use of baseline data and comparative data from the literature and other institutions all contribute to the derivation of standards.

ACKNOWLEDGEMENTS

The researcher would like to extend her gratitude to a number of colleagues and friends who were instrumental in the completion of this thesis.

A special note of appreciatiation is extended to my advisor, Dr. Mary Kennedy, for her inspiration and assistance. Her support, which has spanned coast to coast, was essential to the completion of this thesis that has truly been a distance education effort.

A special gratitude is imparted to the respondents who participated in this study for their cooperation and valuable time.

I would like to convey my appreciation to my colleague Gord Curry who helped me to begin this journey and to my friend Lorna Ebadi who helped me to finish it.

Heartfelt thanks are extended to my sister for her support and the many hours she spent listening to my thoughts, questions and ideas.

Words cannot express my gratitude to my husband, Randy, and son, Nigel, for their love, patience and the many hours

of precious time together they sacrificed while this thesis was written.

Finally, I would like to dedicate this manuscript to my parents whose support and encouragement, not only during this endeavour but throughout my life, has brought me to where I am today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

Nature of the Study

Introduction

The purpose of this study was to explore the program evaluation methodology employed at an autonomous distance education institution, focusing in particular on the derivation and application of standards within the evaluation process. The data gathered was examined to determine their relationship to theory and models in the evaluation literature.

Background Information

This study explored program evaluation methodology at a distance education institution. This institution is an autonomous distance education institution, as defined by Keegan (1986), offering courses exclusively through distance means. It is modeled on the Open University of Great Britain and employs a team approach to course development and evaluation. Although it was originally a conventional university it was rechartered as an autonomous distance education institution in 1975. It enrolls approximately 10,000 students annually, which is the minimum enrollment

deemed necessary economically for a university using a team approach for course development. The institution has also been the site of a large amount of research on the presentation components within distance education (Shale, 1984). The research component of the university has focused on areas such as learner tracking (Coldeway, 1980), behavioral self-control packages (Powell and Coldeway, 1980), seminars as an instructional strategy (Peruniak, 1980), computer generated schedules (Spencer, 1980), peer tutoring (Coldeway, 1980), pacing conditions (Crawford, 1981), and learner motivation (Coldeway, 1980). The efforts to further knowledge of how to maximize student learning and the overall effectiveness of distance education, combined with the institution's success in remaining a viable university since 1975, made it a suitable choice as the subject of this study. For ethical reasons the name of this institution has not been disclosed.

Significance of the Study

Considerable efforts have been made to evaluate distance education (Holmberg, 1986). However, the standards employed for making judgments in the evaluation process are often unspecified (Keegan and Rumble, 1982; McAnany, 1982; Birgitta, 1984; Keegan, 1986). The importance of standards

for making judgments has been acknowledged by authors such as Stake (1975), Wedemeyer (1981), Guba and Lincoln (1981), and Thorpe (1988). However, a lack of linkage between evaluation theory and practice in general has also been acknowledged (Lewey, 1983; Lipsey and Pollard, 1989).

This study is significant both practically and theoretically. Practically, it provides baseline information on the derivation and application of standards within the evaluation process at an autonomous distance education institution. Theoretically it expands our knowledge of evaluation by examining whether there is a link between the evaluation literature and actual practice in evaluation at a distance education institution.

Limitations of the Study

This study was limited by the following:

a) it examined only one institution's approach to deriving evaluation standards;

b) it examined the evaluation of distance education efforts of an autonomous institution only;

c) it focused only on higher education courses and programs offered at the college or university level.

Definition of Terms

For the purposes of this study the following terms and definitions apply.

Distance Education Courses and Programs

In accordance with Verduin and Clark (1991) distance education courses and programs are those in which the teacher and the learner are separated for the majority of the instructional process, and educational technology is used to unite them, to provide two-way communication between the institution and learner and to carry course content.

Conventional Distance Education Institutions

Those institutions that provide both on-campus and distance education courses and programs (Rumble and Harry, 1982; Keegan, 1986).

Autonomous Distance Education Institutions

Those institutions that provide courses exclusively through distance means (Rumble and Harry, 1982; Keegan, 1986).

Evaluation

The process of "examining and judging the value,

quality, or significance of something" (De Machado and Machado, 1991, p.146).

Standards

"A degree of quality or level of achievement regarded as desirable or necessary for some purpose" (Concise Oxford Dictionary, 1988).

Evaluation Standards

"Includes indicators of success or merit" which describe the ideal state of a course or program (Scriven, 1991, p.111).

Evaluation Criteria

Specific measures which indicate whether a given standard has been achieved.

Organization of the Study

The study is organized as follows:

1. Chapter One presents the introduction to the study, including its significance and limitations.
2. Chapter Two presents a review of the literature on distance education, program evaluation, and quality management models and the role of

standards in these evaluative approaches.

3. Chapter Three presents the methodology employed in the study, specifying the data collection process and analysis of results.
4. Chapter Four presents a narrative report of the results of the study followed by an analysis of the data collected.
5. Chapter Five presents the conclusions and recommendations derived from the study.

CHAPTER TWO

Review of the Literature

Distance Learning in Higher Education

Learning at a distance evolved from the correspondence movement in Europe and the United States to the concept of distance education as we know it today. The earliest endeavors date back to the establishment of Isaac Pitman's correspondence school in Bath, England in 1840. This correspondence school for shorthand was the first to grade assigned work and employ the postal system as a means of communication between instructor and student (Verduin and Clark, 1991).

Since then, distance education has grown immensely both in terms of the magnitude of students learning at a distance and the number of universities offering distance education courses. In 1988 it was estimated by the International Council for Distance Education that ten million people worldwide were taking university distance education courses (Kaye, 1988). In 1984, Perry surveyed fifty-two countries regarding distance education offerings for college credit, and found one hundred and forty-two associate or bachelor degrees and sixty-one post-graduate degree programs (Verduin

and Clark, 1991).

Emerging technologies are continuing to make distance education a viable alternative to on-campus courses. In 1991, Smith, in his report of the Commission on Canadian University Education, referred to the expertise of Canadian universities in distance education (Kirby, 1993). Miller and Clark (1994) state: "there is no doubt that technology-based distance learning in higher education and in business expands America's capacity to provide educational opportunities for all citizens and to better prepare a workforce for the twenty-first century" (p. 196). This is an echo of Kirby (1993) who describes the potential that distance education holds in meeting the university's needs for student access in tight budgetary times. Distance learning in higher education has been firmly established as an alternative for students pursuing post-secondary diplomas and degrees.

Keegan (1986) put forth five criteria to define the concept of distance learning which include:

1. The quasi-permanent separation of teacher and learner throughout the length of the learning process: this distinguishes it from face-to-face instruction.

2. The influence of a formal educational organization in planning and preparation of learning materials and in the provision of student support services: this distinguishes it from private study and teach-oneself programs.
3. The use of technological media - print, audio, video, and/or computer to unite teacher and learner and to carry the content.
4. The provision of two-way communication so that the learner may benefit from or even initiate dialogue: this distinguishes it from other uses of technology in education.
5. The semi-permanent separation of the learning group throughout the length of the learning process so that people are often taught as individuals and not in groups, with the possibility of occasional live or electronic meetings for both didactic and socialization purposes(p.37).

Within the general definition of distance education several classifications of this type of education exist. The differentiation between autonomous distance education institutions and hybrid institutions has been advocated by

Rumble and Harry (1982) and Keegan (1986). According to their definition, autonomous distance education institutions are schools or open universities that teach exclusively through distance education means, and hybrid institutions are conventional schools or universities that offer distance education courses through independent divisions (that is, offering both on-campus and distance courses).

Autonomous Distance Education Institutions

In addition to distance education courses offered by conventional institutions, many universities have been established which specialize in and offer courses only by distance means. Moore (1991) lists some of the national autonomous universities that have been established worldwide. Included are:

AL Quds Open University, Jordan
Allama Iqbal Open University, Pakistan
Everyman's University, Israel
FernUniversitat, West Germany
Kyongi Open University, Korea
National Open University, Taiwan
Open Universiteit, the Netherlands
Sri Lanka, Institute of Distance Education

University of the Air, Japan

Universidad Nacional Abrierta, Venezuela

Universidade Aberta, Portugal

Universitas Terbuka, Indonesia

Furthermore, Ilyin (1983) reports that there are fourteen distance education universities in what was then known as the U.S.S.R..

The origin of autonomous distance education institutions is traced back to the establishment of the Open University in Britain (Kaye and Rumble, 1981). According to Perry (1990) the development of the British Open University arose from three post-war conditions. These included the recognition of a requirement for providing education to adults, equal access to education, and the expansion of televised education. Only one decade after its establishment the Open University had enrolled over 60,000 students annually and by 1984 nearly 70,000 students had earned undergraduate degrees from that institution (Rumble,1986).

Following the establishment of the British Open University a number of institutions emerged in Canada such as the Tele-Universite of Quebec, Athabasca University and the Open Learning Institute of British Columbia. The

British Open University, as a well established autonomous institution, served as a model for distance education around the world. A key element of the British Open University and other institutions that operate on similar lines, such as Athabasca University, is the course team approach employed for developing courses (Knapper and Cropley, 1985). Athabasca University offers courses in three faculties which include Arts, Science and Administration. Students who are enrolled in courses are assigned a telephone tutor and have access to library services. Students generally have six months to complete a three credit course and can work at their own rate unless seminars or similar support are integral to the program.

Evaluation Models and Approaches

House (1980) provides an insightful taxonomy of the major evaluation frameworks which consist of eight approaches. The earliest approach to evaluation described by House is termed the "Behavioral Objectives (or goal-based) Approach. As the title implies, it is the objectives or goals that are stated for the program that serve as the sole source of standards and criteria for the evaluation of a program. This approach was initially advocated by Tyler (1950) and further developed by Mager (1962), Bloom (1956,

1971) and Popham (1975). The Objectives Approach was the predominant, and in fact the only, program evaluation method for approximately three decades. According to Guba and Lincoln (1982) it was the advent of Sputnik in 1957 and the criticism of school practices that initiated challenges to this evaluation approach, resulting in a variety of new approaches.

The Systems Analysis Approach focuses on measuring the outputs of education quantitatively and "tries to relate differences in programs or policies to variations in the indicators" (House, 1980, p. 22). The major underpinnings of this approach include the use of test scores and experimental design in an attempt to be as objective as possible when evaluating social or government programs. A thorough systematic evaluation, according to House (1980) will assess "program planning, program monitoring, impact assessment and economic efficiency" (p. 26). The systems analysis approach is employed heavily in government.

The Decision Making Approach to evaluation focuses on providing all the information required, in a given situation, to allow decision makers to choose one of a variety of options that could be pursued. Stufflebeam

(1973) advocated this approach and states "evaluation is the process of delineating, obtaining and providing useful information for judging decision alternatives" (p. 129). Guttentag (1973) provided a quantitative variation of this model while others have taken a more personal approach to determining who the decision makers are, and what information they require to make decisions on a particular issue. House (1980) concludes that this approach attempts to increase the utility of evaluation information. However, who the decision makers are and how the decisions are made varies from context to context.

The Goal Free Approach responded to goal-based evaluation and rests heavily on Scriven's (1973) notion of unintended effects. That is, the evaluator should assess all effects of a program rather than just the program intents, or objectives. House (1980) suggests that the basic underpinnings of this approach include reducing bias by being unaware of program goals and staff desires, and focusing instead on needs assessment to determine consumer requirements from which the program can be evaluated.

A major proponent of the Art Criticism Approach was Eisner (1979) who states that "connoisseurship is the art of appreciation, whereas criticism is the art of disclosure"

(p. 193). According to Eisner (1979), in this approach to evaluation "criticism is the art of disclosing the qualities of events or objects that connoisseurship perceives" (p. 197). To employ this approach, Eisner addresses the need for expertise so that one will be able to illuminate and expand the subject under study. House (1980) suggests that to expand perception is the goal. From a critic's writing the reader will know the successes and/or shortcomings of a program. However, more importantly they will have an expanded perception of the program, which is the ultimate goal.

The Professional Review (Accreditation) Approach is described by House (1980) as an evaluation framework in which schools from elementary to the professional (law, medicine) are evaluated or reviewed by their peers. In many instances this type of review is conducted for accreditation.

As an example of the Professional Review, House (1980) describes the National Study of School Evaluation (1978). This study publishes both service-specific and curriculum-specific criteria for evaluating programs. Each criterion is then evaluated on a point scale from "missing" to "excellent" (p.195). Initially, the school staff evaluates

their school using these criteria. Then the accreditation body validates the results by conducting their own evaluation using the same criteria. Any differences are noted, recommendations are made, and accreditation may be awarded.

Accreditation can also be conducted for individual departments within an institution. An example of this approach is the Council On Program Evaluation (COPE) at the University of Illinois (House, 1980). In this type of scenario departmental staff may be interviewed or asked to respond to a questionnaire on specific issues. Based on the data collected the evaluation committee or council responds with recommendations for improvement.

House (1980) suggests that although the evaluation or review processes may not be identical in all cases, most schools are conducting them. He points out that a major difference in professional schools is that the criteria for evaluation are based on the judgment of the professional personnel who sit on the committee. Most often committee members consist of peers within the same profession.

The Quasi-Legal (Adversary) Approach can be compared to the use of commissions or panels of investigation who hear evidence about an issue and draw conclusions from it. The

commissions simulate the legal process although they are not held in a court of law. According to House (1980), this type of legal adversary process has been employed like a hearing to evaluate programs. He cites two mock trials that were held in Hawaii and Indiana University, where opposing teams presented arguments to either judges or a jury, to evaluate what course of action should be taken.

Wolf (1975) advocates the use of rules and procedures in this approach and suggests "the judicial approach provides for the structured consideration of alternative arguments and inferences to keep the evaluation both intellectually honest and fair" (p.185). House (1980) also concludes that several types of legal processes can be applied to evaluation, however he also suggests:

> the legitimacy of the approach depends on whether one accepts the particular procedures employed, the hearing officer, and the deciding panel or jury. In the common law itself, court action is usually decided by judges alone (p.39).

The final evaluation approach described by House (1980) is the Case Study (or Transaction) Approach. The purpose of this approach is to develop a better understanding of a

program through primarily qualitative research methods such as interviews and observation.

Stake (1975) advocates the Case Study approach. In his Responsive Model he states:

> To do a responsive evaluation, the evaluator conceives of a plan of observations and negotiations... He finds out what is of value to his audiences, and gathers expressions of worth from various individuals whose points of view differ. Of course, he checks the quality of his records; he gets program personnel to react to the accuracy of his portrayals, and audience members to react to the relevance of his findings (p.14).

House (1980) describes three other similar approaches within this framework. This includes democratic evaluation which is described by MacDonald (1974) as "an information service to the community about the characteristics of an educational program..." in which "the main activity is the collection of definitions of, and reactions to, the program" (p.226-227). Similarly, Parlett and Hamilton (1977) advocate "illuminative" evaluation and suggest it "aims to discover and document what it is like to be participating in

the scheme, whether as a teacher or pupil; and in addition, to discern and discuss the innovation's most significant features, returning concomitants and critical processes" (p.19). Guba's (1978) naturalistic evaluation is also identified by House (1980) as being encompassed within this framework.

House (1980) suggests that, overall, the case study eases the evaluator's burden by "attempting to represent all significant value positions within the case study, drawing its criteria and standards from those positions and letting the reader of the study weigh and balance these elements with his/her own mind" (p.42).

Total Quality Management

The quest to evaluate or measure the success of institutions is also addressed in a variety of initiatives focusing on quality. Total quality management (TQM) is one approach, which focuses on continuous improvement within an organization to meet customer requirements. Quality improvement within this model is defined by Harris and Baggett (1992) as "continuous improvement of processes in a cycle of plan, do/check (assess), act (on the assessment and then continue to), plan-do-check-act" (p.23). The ideology behind TQM is the notion of examining the quality of each

aspect of an activity and working constantly to refine it even if the improvement that occurs is in small, gradual increments.

TQM has made significant inroads into institutions of higher education in recent years. Van Vught and Westerheijden (1994) suggest the increased attention to this concept is due to expanding enrollments, financial constraints and the integration of technology in society. The application of TQM to higher education is advocated by Sevick, 1993; Schmoker and Wilson, 1993; Cross, 1993; Sutcliffe and Pollock, 1992 among others, who contend that institutions can consider their service as a product, employ customers' criteria to judge quality, and strive to improve existing standards. The application of TQM to higher education has been explored in depth by several authors. The analysis provided by Capezio and Morehouse (1993) provides insight into the components that form the basis of this management model.

Capezio and Morehouse (1993) contend that measurements are integral to determining how well a company is meeting customer requirements. To accomplish this task, several types of measurements are discussed. Determining baseline measurements is cited by Capezio and Morehouse (1993) as particularly critical as they "respond precisely to customer

requirements and set standards of excellence throughout the organization..." (p.186). Baseline measurements can be taken for a variety of areas such as the quality of products, and processes, and the amount of output and time involved, depending on the organization.

Baselines provide a vision of excellence enabling goals for improvement to be made. Goals for improvement are developed in relation to baselines, by bringing together and examining customer requirements and organizational goals.

Benchmarking is a second type of measurement, which involves examining the practices and standards of another department or company, and determining how they compare to one's own. For example, what are their baselines or targets for improvement.

To determine customer requirements Capezio and Morehouse (1993) suggest two techniques. They contend the quality dimension process "engages people familiar with the needs of the customer and the product" to develop definitions and examples of dimensions of quality (p.152). These should be specific statements including adjectives, behaviors and the like that reflect customer requirements.

The second, the critical incidents approach, entails asking customers about an either good or bad experience. This information is generated through customer interviews

and provides specific customer requirements, categorized into similar groupings. The use of questionnaires are advocated to find out how the organization or its specific aspects is meeting customer needs.

The most common equation cited by Capezio and Morehouse (1993) to measure deficiency is "quality is equal to the frequency of deficiencies in a particular activity divided by the opportunity for deficiencies" (p.189). One derives a percentage of deficiency in relation to customer requirements, from which an incremental target for improvement can be set from the baseline standard, in the continuous improvement process. While Capezio and Morehouse do not employ the terms standards or criteria in their description, the relationship between setting some target for improvement and the use of evaluation is clearly established. According to Capezio and Morehouse (1993) evaluation is in fact ongoing in the TQM process. They state "evaluation is considered during the planning and design stages of each improvement strategy when criteria for success are established in the objectives. From the beginning, people know how they will evaluate the new initiative" (p.261).

Other Quality Initiatives in Higher Education

Performance indicators, according to Adams (1991) are the management ratios which allow for measurement to be pursued. The focus on measurement and audiences in evaluation is inherent to the approach of employing indicators in higher education that she describes. Adams (1991) identifies the group examining performance in higher education as either providers or clients. The providers are defined as those "who supply funding, direction and expertise to the system..." and include funding agencies (national and local), "college governors and managers" and the instructional body (p.5). Clients are all customers of the system, directly or indirectly, such as parents, students, employers, and so forth. Altogether, these groups are the stakeholders and it is the indicators that they deem valuable that are employed in an evaluation. Indicators can be focused on a particular context or can be applied to more than one context Employing indicators such as student-staff ratio, which can be applied in more than one context, is advocated in order to maximize effectiveness.

Adams (1991) provides examples of indicators for different stakeholder groups. She suggests that provider-efficiency indicators include administrative (non-instructional) costs per student, use of facilities, course

completion ratios, actual and ideal enrollment ratios, student-staff ratios, and comparison of current students to those already qualified. Consumers may consider other indicators as pertinent such as student access to teachers and the length of a course.

Effectiveness indicators of providers are described by Adams (1991) as including the satisfaction of the customer and the "entry-exit level of competence of the student" (p.14). For consumer groups the focus shifts to meeting the needs of, or satisfying students.

Adams (1991) outlines a model for managerial evaluation which consists of thirteen procedures. The procedures include: identifying stakeholders for the evaluation; determine their concerns; write the concerns into objectives; define the focus and limitations of the evaluation based on these objectives, issues and concerns; select appropriate indicators from lists of possible indicators; relate indicators specifically to objectives, issues and concerns; determine data sources; define data collection methods and time line; define methodology for the data analysis; determine outputs resulting from the evaluation; identify how the outputs will be presented and to what parties; and, "review the cycle of managerial evaluation and incorporate outputs into planning processes

for the next development phase" (p.11). These 13 steps are practically identical to the procedures outlined in Stake's (1975) responsive model.

Although the term performance indicators is employed in the evaluation of higher education described by Adams (1991), the similarity to responsive evaluation and to TQM is evident, given the notion of deriving standards and criteria based on the concerns or requirements of stakeholder groups.

Evaluation Approaches in Distance Education

Gooler (1979) offers several criteria which are deemed especially important when evaluating distance education. These include: promotion of equal opportunity to prospective students; accessibility from varying locales; meeting needs of the population; quality of courses compared to the traditional; successful course completion; contribution to our knowledge of learning; impact of distance education on students, society and institutions; and cost-effectiveness.

Paul (1990) contends that in order to measure student success in an open learning environment, institutions must "develop clear criteria of success and tangible indicators with which to measure them..." (p.77). He addresses several indicators deemed useful for distance education evaluation

including:

a. completion rates

b graduation rates

c. persistence rates (taking another course)

d. "cost efficiency and effectiveness: the cost per course, per completion, and per graduate.

e. skill development: the degree to which students develop their independent learning skills so that they can increasingly take responsibility for their own learning.

f. post-graduation performance: in subsequent education and employment" (p.78).

The need for measurement in distance education evaluation is also addressed in other models. Keegan and Rumble (1982) recommend a four part evaluation plan to determine the overall effectiveness of a distance education institution. In order to evaluate performance, they contend that the following must be established: the quantity, status, relative cost and quality of the learning attained. Keegan and Rumble (1982) discuss the use of standards derived from the university's aims and goals for criterion-referenced evaluations. They employ a norm-based approach to conducting appraisals of several institutions.

Within Gooler's (1979) criteria, Verduin and Clark (1991) suggest that "sub-goals or activities that are more precise and measurable, that will give direction to program development" should be specified (p.189). They contend that these sub-goals are the "value expressions to which the evaluation process is applied" (p.189). For example, in terms of access, an example cited as a sub-goal is to have resources available to students at all times. In evaluation terminology, access would be the standard, and resources available to students at all times would be one criterion measure of that standard.

Rumble (1986) offers an approach to evaluating course and program effectiveness based in part on Birch and Latcham's (1985) model. This model is described by Rumble as requiring that "critical outcomes are identified, satisfactory levels of performance are set, and data definitions and rules are applied. Effectiveness is based on space available, student enrollment, course completion, student passes, and graduate jobs. Efficiency is determined by examining the money expended per student or graduate. This model does address the issue of standards and criteria for evaluation, in setting satisfactory levels of performance through ratios (criteria) for each critical outcome (standard).

Distance Education Evaluation Applications

Markwood and Johnstone (1992) suggest that distance education programs evolve in three areas to function effectively within an organization. These include "technological reliability" in terms of accessibility, "institutional support" for students, faculty and administrative personnel, and "organizational design and development" focusing on the integration of distance education into the establishment. Wagner (1993) conducted focus groups on distance education and found that "anticipating and supporting the needs of the users of technologies, programs and services emerge as being critical to the success of distance educational enterprises" (p.28). Harrison, Seeman, Behm, Molase and Williams (1991) lend further support to the importance of these factors in distance education. They identify "instruction, management, telecommuting and support" as critical to the effectiveness of distance education (p.65).

The factors described as critical for success above are closely aligned with standards and related criteria that are employed for the evaluation of distance education programs. Perry (1977) describes evaluation at the British Open University (BOU) as employing weekly student questionnaires on aspects of teaching materials such as: "time on each

unit; course material that was interesting and essential", and non-essential material (p.261). In addition information on the availability of equipment and hours when students could make use of it, and feedback to tutors from students is also collected on a continuing basis.

Hotchkis and Driedger (1990) state that at Athabasca University educational success is "measured by the number of students who complete a course; the number of students who successfully graduate; and the number of students who take subsequent courses" (p.41). They also suggest that examining students' lives on a day-to-day basis in considerable qualitative depth provides insight into student difficulties.

The focus on feedback from students and enrollment/graduation type data collection is evident in the examination of evaluation practice in autonomous distance education. A myriad of standards and criteria are employed with a great deal of similarity from one institution to the next. Bartels (1991) describes evaluation by the Center for the Development of Distance Education at the Fernuniversitat as focusing on the evaluation of courses to improve course materials, and the evaluation of the system (i.e. motivation, drop-out, access to computers, learning conditions, gender, and exam participation).

Bartels (1991) states that when evaluating new and revised materials course critiques are standard procedure, and are employed as a basis for further evaluation. The course critiques are specific and focus on the quality of course units in terms of text comprehension, use of learning aids, presentation of content, assignments and learning objectives. A course evaluation coordinator collects all information from course critiques and summarizes it for authors.

Group interview guides have also been developed as a tool for evaluation. Bartels (1991) describes them as addressing the presentation and content of material; assignments and tests; "use of learning aids and literature; time needed to complete work; and organizational problems" (p.62).

Overall, Bartels (1991) suggests that there is more of a focus on system rather than course evaluation. System evaluation examines drop outs, student motivation, graduates' experience in and out of school, and exam participation as well as successful students who transfer or just drop out.

Ganor (1991) describes ongoing evaluation at the Open University of Israel as including the examination of students in relation to progression and background

characteristics; computer and teacher graded course work; and final exams in relation to course work. In addition survey research is employed periodically on course materials, method of instruction, student population in relation to persistence and success, staff in relation to training and roles.

Ganor (1991) describes the evaluation of courses as focusing specifically on the validity of assignments and final exams; study material including:

> suitability for self study, degree of difficulty, interest, clarity of texts, graphic presentation, audio-visual aids and updating of material; and, tutoring to include the content and structure of tutorial meetings, teaching strategy, student attendance, individual tutoring, marking assignments, and experiments with new ways of tutoring (p.87).

The evaluation of programs at Indira Gandhi National Open University is described by Koul (1991) as focusing on the planning of a course, the quality of course materials and student services. The planning of a course is evaluated in terms of: "need for it; utility of the course; adequacy of content and media; and economic viability" (p.94). Koul

provides examples of criteria employed for economic viability such as money spent on meetings, course author orientation, support services, equipment, materials, and the cost of learning for students.

Koul (1991) describes criteria employed for evaluating course materials as including whether content is adequate; aimed at the level of the student; self-instructional; employed with appropriate media; relevant; accessible; helping the learner to pass; and useful in relation to what other universities are using.

According to Koul (1991), the evaluation of student services is examined through four categories or criteria. The first is tutor evaluation which addresses: communication between tutors and students; comments written on course work; the reliability and validity of grading; and turn-around time for course work. The second component examines in-person sessions for timeliness, rapport, counselor motivation and students' response. Support services are examined in terms of the quantity and quality of services provided. The final component addresses the evaluation of the reliability and validity of course work in relation to final exams.

Koul (1991) addresses a second tier of evaluation at the Indira Gandhi National Open University. This consists

of the evaluation of selected pilot courses, special evaluation due to a problem or new design/method, and routine evaluation of a course in its first year, and of assignments and final exams.

De Machado and Machado (1991) state that evaluation at the National Open University of Venezuela is carried out by members of the administration and academic faculty. They function to derive standards through discussion and work together professionally as a committee. There are two committees. One conducts the evaluation of administration and the other of instruction.

The evaluation of instruction is described by De Machado and Machado as consisting of the validation of materials before using them; the evaluation of the teaching learning process; and evaluation of students, including course evaluation and self-evaluation. The evaluation of administration examines services, personnel, and budgets/cost.

Woodley (1991) defines program evaluation at the British Open University (BOU) as focusing on "how well a particular educational programme, curriculum, or teaching method works, how it might be improved, and how it compares with alternatives" (p.208). Woodley (1991) describes evaluation activities carried out in formative and summative

evaluation. Formative evaluation consists of: critical commenting, referring to peer review draft materials and critical commenting on materials by teachers and writers; developmental testing, referring to student testing materials; doing assignments and tests; attending seminars, and commenting on materials on questionnaires and in interviews.

Summative evaluation consists of primarily student and tutor feedback. Woodley (1991) describes feedback from tutors as focusing on student problems and materials, and feedback from students which can occur both during and after the course. Student feedback consists of information on what content they have used, how they rate instruction, the presentation of material or course components, and problems they have had with concepts or ideas in the content.

Woodley (1991) also expounds on various other types of evaluation conducted at the BOU. This consists of cross-sectional evaluation of components such as access or the use of resources, developmental evaluation of what students learn and overall system evaluation.

The review of autonomous distance education evaluation practice indicates that there is generally some form of on-going and/or formative evaluation of course materials as well as summative evaluations. It is important to note that

although criteria for standards have been identified, they are not stated in a measurable form. Overall, the evaluation procedures are still heavily focused on student feedback. Furthermore, with the exception of the National Open University of Venezuela, there is no indication of how the standards and criteria for evaluation are derived.

Problems in Distance Education Evaluation

Despite the success of the British Open University and the proliferation of distance education programs worldwide, most autonomous distance education institutions are still judged to be inferior to conventional institutions by conventional institution faculty and administration (Holmberg, 1989). Thorpe (1988) notes that there are a paucity of examples of evaluation of distance education outside of those conducted at autonomous distance education universities. She suggests that there is little evaluation done in conventional universities. A need to establish the value of distance education exists due to the relatively new approaches to education employed at these autonomous institutions. Thorpe (1988) argues that evaluation of open learning is needed to determine program/course effectiveness, to improve the quality of learning and to present to external bodies when necessary. This is

consistent with the concerns of the University of Alaska (1990), where the administration suggests that evaluation is critical to ensuring program effectiveness due to the diversity of students, relatively low level of interaction, and variety of locations from which students learn in distance education.

While there is evidence that considerable efforts to evaluate distance education have been made (Keegan, 1986; Thorpe, 1988; Miller and Clouse, 1994), there are difficulties inherent in the evaluation of distance learning (Evans, 1985). On a basic level, as students are unavailable, evaluation cannot take place on a day to day basis. In addition, studies suggest that student and expert opinions and attitudes appear to be a large part of distance education evaluation.

McAnany (1982) advocates that evaluation focus more on factors external to the distance education program that affect student success. This ascribes to Scriven's (1972) notion of evaluating unintended effects. Since his development of the goal-free model, many evaluation approaches have emphasized the need to look beyond program goals. McAnany (1982) concludes that outside of extensive evaluation at the British Open University, little data exists beyond enrollment and budget data, in most open

universities. The focus on student feedback and limited outcomes can result in a neglect of many elements that should be part of a comprehensive evaluation. The need for a comprehensive approach to evaluation is addressed by many authors including Patton (1982) and Lincoln and Guba (1985).

While course completion versus attrition rates have also been included in many evaluation studies, Coldeway and Spencer (1980) identify two major difficulties in carrying out this process. They are: defining which students are enrolled, and ascertaining when to tabulate student completion once the course has begun. Lack of precise measurement on these factors also results in problems with comparative studies of distance education institutions. Similarly, Guba and Lincoln (1982) have suggested that comparative studies such as those based on learner outcomes have outlived their usefulness.

Paul (1993) addresses difficulties in defining what is success in distance education. For course completion rates, one problem lies in the fact that students may obtain what they want from a course without completing it. In addition, attempts to derive a standard as such from similar institutions can prove hazardous depending on how course completion rates are calculated. Shale (1982) addresses this issue extensively by contrasting the difference in

course completion rates at the BOU, when non-starters or early withdrawals are included in the calculation. For example, he cites a completion rate of 70% at the BOU as being lowered to 36% if the non-starters were included. Furthermore, he found that of the 1978-79 student intake at Athabasca University, 50% of the students fell into the non-starter category. The impact of such difficulties on measurements of course completion, graduation and persistence rates, as well as cost, cannot be denied.

Paul (1993) argues that a criteria of skill development in terms of "producing independent, self-directed learners" (p.82) is a more relevant measure of success. He suggests that Distance Education institutions must increase their ability to produce independent learners, although he acknowledges that this criterion is vastly more difficult to measure than the typical graduation/completion rates. Although Paul (1993) does not provide any way to measure this criterion, he cites Moore's (1986) notion of "training tutors and course writers in self directed learning, offering student support services on a demand basis and decoupling the teaching function from the accreditation function" as a starting point (p.94).

Coldeway (1986) further expounds on the difficulty in defining criteria for success in distance education. He

suggests a variety of factors such as differences in institutional, faculty and student notions of success. Coldeway cites a thirty percent course completion rate at Athabasca University. He examines students' perceptions of success in terms of course completion (yes/no) versus satisfaction with the course/program, and in terms of high/low grades versus course/program satisfaction. From his data analysis it can be concluded that if the number of students that withdraw or achieve low marks, but are satisfied, is higher than those in the low marks/withdrawal category that are dissatisfied, then this may indicate a successful course, depending on the criteria for success.

Most distance education studies have not addressed the derivation and application of standards in the evaluation process (Birgitta, 1984; Keegan and Rumble, 1982; McAnany, 1982; Keegan, 1986, p.251). Daniel and Snowden (1981, p.224) discuss the importance of evaluation in small open universities, but contend that although standards are needed because of the relative newness of distance education, there is little to base them on. Guba and Lincoln (1982) suggest that without standards there can be no judgment. This view has also been expressed by Stake (1975) in his responsive model and by Weiss (1991), who advocates the use of "operative goals" as the standards upon which a program is

evaluated.

Scriven (1991) states:

> the evaluation process normally involves some identification of relevant standards of merit, worth or value, some investigation of the performance evaluands on these standards, and some integration or synthesis of the results to achieve an overall evaluation or a set of associated evaluations (p.139).

He suggests that evaluation consists of two arms: one which gathers data; and, the other that "collects, clarifies, and verifies relevant values and standards" (p.5). Scriven's (1991) notion of the two arms of evaluation is depicted in Figure 1.

Scriven's

Evaluation Framework

ARM A	ARM B
GATHER DATA FROM PROGRAM IMPLEMENTATION	GATHER DATA FROM LITERATURE, AUDIENCES, AND DOCUMENTS
ANALYZE AND SUMMARIZE DATA	CREATE STANDARDS AND CRITERIA/INDICATORS
COMPARE INFORMATION & MAKE JUDGEMENTS	

Figure 1

It is the evaluation data collection illustrated in Arm B of Figure 1 that serves as the focus of this study. Scriven (1991) refers to this as the values side of evaluation and suggests several steps for clarifying them. Included is the determination and removal of "inconsistencies in individual sets of values; misunderstandings and misrepresentation of values and false factual assumptions underlying them" (p.5). Scriven (1991)

further suggests that the difference between needs and wants must be determined and problems such as ascertaining that all "relevant dimensions of merit" have been identified; deriving methods for measuring them; and, "weighting the dimensions in some way that accurately reflects our intentions"; are addressed (p.5). Scriven (1991) concludes that the resulting standards must be validated and that when data from the program have been collected, the two arms are then synthesized so that judgments can be made.

Standards

Standards have been examined in-depth within the field of measurement. Within this framework standards are described by Cizek (1993) as the derivation, description, and employment of rules which will facilitate or enable one to make judgments. This notion is consistent with general dictionary definitions of standards as "a model to be followed or imitated, established by authority or by general consent, a degree of quality, level of achievement etceteras, regarded as desirable or necessary for some purpose" (Danby, 1989). In terms of evaluation, the Concise Oxford Dictionary (1988) provides a definition of a standard as "a degree of quality or level of achievement regarded as desirable or necessary for some purpose".

The difficulty that can occur in attempting to define standards for the purpose of evaluation has been addressed by several authors. Pring (1992) states "the difficulty in talking about standards is that the concept is like truth or beauty, both logically indispensable and yet impossible to define without considerable philosophical elaboration" (p.21). Harvey and Green (1993) suggest that standards for assessing quality are subjective and variable to changing needs. They advocate a stakeholder approach which includes "students, staff, accreditors, accessors, employers and government" to determine what quality means (p.144). Guba and Lincoln (1981) acknowledge that gathering the information required to derive standards requires that the evaluator be quite knowledgeable in data collection and analysis procedures. This is an echo of Stake and Denney (1969) who comment on the daunting task of developing standards due to problems in retrieving information from experts and the professional literature. Furthermore, in terms of distance education, Paul (1993) addresses the difficulty of defining criteria to measure success.

De Machado and Machado (1991) define evaluation as the process of "examining and judging the value, quality, significance, amount, degree or condition of something" (p.146). Therefore, they contend that "evaluation is a

process which implies a comparison"... of an object to another that is used as a "standard of comparison" (p.147). These standards are described as:

> defining an ideal state, an acceptable or anticipated behavior, an intended result, or goal which, in turn, implies the need for collecting relevant information on the exact state of the object for evaluation and the criterion (ideal state to be used for comparison) (p.147).

Scriven (1991) defines the relationship between standards and criteria more precisely, stating that in evaluation the notion of standard is employed "to include indicators of success or merit; primary indicators are tied directly to standards rather than loosely related or a secondary outcome" (p.111). In this sense, the standards function to describe the ideal state of a program and the criteria are employed to state specific measures which indicate whether a given standard has been achieved. For a program it is expected that many standards of the ideal state will exist, and for each standard several criteria will exist to measure the extent to which the standard is being met.

The focus on judgment as inherent to evaluation and the

need for standards is evidenced throughout the history of evaluation. Nevo and Friedman (1992) conducted a study of evaluation reports and found that judgment definitions were more common among academic evaluators and in summative evaluations. Passow (1987) addresses the importance of standards indirectly in his discussion of evaluation reporting procedures. He states that an evaluation report "should contain sufficient information to respond to the concerns and issues of the audiences and provide enough information so that the judgments upon which the recommendations are based are clear". The relevance of standards in evaluation is further evidenced by the number of studies attempting to define how one should set them (Fink and Kosecoff, 1980; Fadale and Winter, 1985).

Despite these efforts the literature is still not extensive and problems have been identified. Although the standards for the evaluation of programs, projects and materials (SEPPM) are identified as useful by Lewey (1983), difficulties are also described due to the lack of theory upon which the standards are based. In a national study of core standards and performance measures in secondary and post-secondary vocational education, McCaslin and Headley (1993) note that the performance measures, standards and assessment methods vary greatly from one institution to the

next.

Standards and Evaluation Theory

Theoretically, the notion of standards has been explored in depth by authors such as Stake (1967-75) and Guba and Lincoln (1981). Guba and Lincoln (1981) suggest that the focus on judgment as a main activity of evaluation in Stake's Countenance model (1967) is valuable. This is due in part to his distinction between absolute and relative standards and his suggestions for how to derive them. Absolute standards are described by Stake as related to personal judgments of individuals or groups, while relative standards are based on comparison to characteristics of other programs. Stake (1967) further contends that the evaluation must determine if a standard has or has not been met prior to making a judgment.

Despite these contributions, Guba and Lincoln (1981) contend that Stake, in his early model, did not specify the manner in which standards were to be derived adequately. In their approach to evaluation, Guba and Lincoln propose strategies for developing standards and making judgments in their expansion of Stake's more recent Responsive Evaluation Model (1975).

Stake (1975) states that an "educational evaluation is

responsive evaluation if it orients more directly to program activities than to program intents, responds to audience requirements for information, and if the different value perspectives present are referred to in reporting the success or failure of the program" (p.14). He contends that "different purposes and information needs of different audiences" (p.14) should guide the focus or aim of an evaluation. The focus of Stake's Responsive Model is on the staff, audience and clients of a program, and it is from them, their issues and concerns, that the standards and criteria are formulated.

This focus on the audience or clients of a program remains inherent to the process in Guba and Lincoln's (1981) naturalistic evaluation approach. They define evaluation as the "process for describing an evaluand and judging its merit and worth" (p.45). Merit is described as relating to the "intrinsic" value of an entity (p.45). Merit is described as absolute when it is determined by assessing to what extent an entity matches standards that have been specified by experts, and relative when comparisons of the entity under study are made to others of a similar type. Worth is the "extrinsic" value of an entity (p.45). It is determined by assessing to what extent an entity or program conforms to a set of criteria or needs as determined by

people who are involved in or affected by that entity. Guba and Lincoln (1981), borrow Stake's term, and label those people who are related to or affected by the evaluland as "stakeholders".

Guba and Lincoln further distinguish between the purpose and source of standards for formative and summative and merit and worth evaluations. The purpose of a formative, merit evaluation is described as functioning to perfect a program currently being designed. The purpose of a summative, merit evaluation is to assess the success of a completed program in terms of professional expert standards, to determine if the program can continue without revision. The source of standards, advocated by Guba and Lincoln (1981) for both formative and summative merit evaluations are some sort of expertise such as professional opinions and literature on the entity being evaluated.

Guba and Lincoln (1981) suggest the purpose of formative worth evaluations is to revise courses currently being designed so that they meet the local requirements. It is suggested that standards for this type of evaluation are derived from an assessment of local needs, values and contexts.

The aim of summative worth evaluations is to examine a completed program to determine if it is meeting local

requirements in its present state, and whether it will continue to operate as such permanently. The source of standards, suggested by Guba and Lincoln, for this type of evaluation is through a needs assessment. The evaluation will then judge if the program is meeting the needs identified.

Standards in Total Quality Management and Quality Approaches

The notion of developing standards or measurements of success in TQM is very much in evidence. Kaufman (1991) describes the application of TQM at South Bank University as involving the development of service quality indicators which are then monitored. The New Zealand Qualifications Authority (1993) contends that procedures such as developing, reviewing, and updating standards, performing self-evaluation and quality assessments are critical to achieving quality in educational/training initiatives . Furthermore, Osbourne (1993) concludes that TQM cannot be incorporated into educational institutions unless standards have first been derived. This is consistent with Duffy (1991) who addresses the ten-item Deming Application Prize for TQM, providing further evidence of the role of standards in the process. Item six on standardization consists of the assessment of the following: "systemization of standards;

methods of establishing standards, revising and abolishing standards; outcomes of the establishment, revision or abolition of standards; contents of the standards; utilization of statistical methods; accumulation of technology; and utilization of standards" (p.42).

Lewis and Smith (1994) state "[that TQM] systems assume that there are standards and/or formats that define how tasks are handed over" (p.284). They acknowledge that this is a critical and difficult part of setting up total quality management because people differ on what the standards should be. They suggest that "fitness for purpose" is a good starting place and that using this definition "means ensuring that all debates on quality are tested against customer expectations" (p.285).

When developing a TQM system in higher education, Lewis and Smith (1994) contend that measurement of output is critical. They suggest that "exams, quizzes, assignments, professor evaluations, surveys of students, alumni, parents, and employees as well as pass rates on professional exams, and admission success rates to graduate schools are indicators of the quality of educational programs at a university" (p.67).

Lewis and Smith (1994) distinguish between internal and external customers in their application of TQM to higher

education. They suggest that internal and direct external customers should be prioritized (in this order) as they directly receive the university's courses of study, services and research efforts. Internal customers are described as including "students, faculty, programs and departments that impact on a program" as well as "employees and units, departments or divisions that influence a service or activity" (p.92). Direct external customers are identified as employers, other schools, and suppliers who receive a student, product or service from the university. Indirect external customers, on the other hand, consist of donors, legislative groups, the city or town served as well as boards or organizations who conduct accreditation.

Other quality initiatives in higher education also focus on the customer or stakeholder and the development of indicators or standards as integral to the process. Tannock (1992) addresses the use of quality assurance methods for higher education that were developed by the Engineering Professors' Conference (EPC).

The approach employed is derived from "British Standard 5750 Quality Systems and total quality management" (p.109). Tannock (1992) describes the quality assurance principles which are employed to develop "quality systems standards" as entailing (p.109):

a. the setting of quality objectives;

b. planning activities to meet these objectives;.

c. documentation of the quality plans;

d. performance of the activities and the collection of data on the quality of performance;

e. the review and assessment of how effectively the activities were performed;

f. identification and correction of any deficiencies.

All of these processes are conducted within the guiding framework of meeting customer needs and improving quality on a continuous basis.

The EPC approach includes four main components. The first, which is identified as the structure of an institution, deems that all functions within it must be examined. Specifically, these include "the central institution, academic units, support and administrative services" (p.113).

The second component consists of the development of a quality mission for the institution overall which can then be specified into objectives for meeting and improving quality at various levels.

The third component addressed is quality systems, which is linked to the development of objectives. This component

of quality focuses on documenting existing systems in each area of the institution, which will serve as a baseline from which improvement initiatives can be implemented and assessed. Quality objectives result from this process. Tannock suggests that, through the documentation process, gaps and anomalies become evident which can then be addressed.

The final component of quality addresses the review and improvement process. This process entails assessing activities against the quality objectives previously set. In addition the quality objectives themselves are assessed against customer requirements. This provides feedback on progress, deficiencies and customer needs (which may change) that is then fed back into the system to establish revised quality objectives. In this manner the notion of continuity is realized, as quality objectives or standards are revised continuously rather than remaining static.

The EPC model addresses many institutional elements. One is the design and provision of programs of study. The stated requirement is:

> The institution together with the academic units shall establish and maintain procedures to plan, assure, and validate the design and provision of programmes of

> study, taking into account all aspects of teaching, learning and assessment (EPC, 1992, p.121).

Tannock (1992) expands on points which must be addressed for each study program. These include the examination of:

a. program purpose in relation to customer requirement;
b. program objectives in relation to what must be achieved by students;
c. instructional/learning strategies and program content, organization and level;
d. student assessment methods in comparison to instructional/learning strategies;
e. facility and resource availability from the program department and support units;
f. program fit in relation to other departmental programs;
g. how the quality of instruction and student assessment will be evaluated.

While individual academic units are responsible to address these points, Tannock (1992) suggests they may employ input from "other institutions, professional bodies, industry or

commerce" as well as from other academic disciplines (p.122).

Standards in Distance Education

It appears that, in autonomous distance education institutions several types of evaluation and standards are integrated in the assessment of the quality of programs. Evaluation types which may impact on courses offered, to some extent, include the evaluation of systems, special projects, and developmental studies. In addition, routine evaluation (which is sometimes used as a basis for further evaluation) of course materials, final exams in relation to course work, computer and teacher-graded assignments, student progress and background characteristics, first year courses and, computer and teacher graded assignments appears to be common.

Formative evaluation focuses on the need, the usefulness, adequacy of content and media, and the economic viability of courses. For each of these standards a number of criteria would be employed to measure whether the standard had been met.

Formative evaluation also includes the validation of course materials before using them in courses, in some cases. This normally entails a review of materials by

peers, authors and instructors, as well as testing through assignments, exams, exercises, and instruction with students.

Summative evaluation occurs at the end of a course and has focused on a number of key standards. Course materials, for example, looks specifically at suitability for self study, difficulty, clarity, interest, graphics, audio visual aids appropriateness, currency (updating), and utility of content. Tutoring, which is also critical to the learning process, looks specifically at structure of meetings, content, teaching strategy, attendance, one on one tutoring, grading of assignments (reliability/validity), initiatives in new methods of tutoring, rapport/communication between tutor and students, written comments on course work, turn around time for corrected course work, student response, counselor motivation, and the timeliness of session.

Support services and cost are also standards. Support services evaluation focuses on both the quantity and quality of availability and access to resources. Cost examines the amount of money spent per course, per completion, and per graduate.

Additional criteria which have been described by authors but not in relation to any particular standard include presentation of materials, pre-course knowledge,

comprehension of materials, learning aids, difficulty experienced with material, time on materials or components, teaching/learning process, assignments, course evaluation scheme, learning objectives, student self-evaluation, tests, instruction, organizational difficulties of the course, persistence rates, drop outs, completion rates, post-graduation success (work or school), development of independent learning, and exam participation.

Inferences From The Literature

The focus on developing standards, based on customer needs, to measure improvement in the TQM and other quality initiatives, can be compared to the derivation of standards for evaluation from stakeholders advocated by Guba and Lincoln (1981) and Stake (1975). Customers in TQM may be internal such as students or external such as donors or accreditors, but are very much all inclusive (Lewis and Smith, 1994). Stakeholders in evaluation are described as including everyone involved in or affected by the program under study (Guba and Lincoln, 1981), and can range from local business people to students.

In the same way that precise statements of customer requirements (Capezio and Morehouse, 1993) are developed in groupings for a particular objective or standard in TQM, in

evaluation specific and measurable criteria are developed to determine if a standard has been met. Just as benchmarking in TQM is employed to derive standards from other expert (successful) institutions (Capezio and Morehouse, 1993), expert opinion and literature is applied to standards derivation in the evaluation of merit (Guba and Lincoln, 1981). In TQM standards are developed from customer needs in relation to baselines. This is comparable to worth evaluation (Guba and Lincoln, 1981), which is focused on whether a program is meeting the needs or criteria determined by stakeholders. Furthermore, Tannock expands on points to examine in the quality assessment of programs which correspond to standards for the evaluation of courses/programs.

While the terminology for measuring success can range from standards and criteria, objectives and criteria for success, to primary and secondary indicators in the various models, the basic underpinnings are similar. Therefore, despite the difficulty in the standard setting process it is concluded that a number of procedures are either theoretically advocated or practically applied.

Theoretically, the derivation of standards is differentiated based on the purpose of an evaluation. Guba and Lincoln's (1981) naturalistic model suggests that merit

evaluations derive standards from some form of expertise or similar expert programs. Professional opinion and literature serve as the source for deriving standards when perfecting a course (formative) or determining if it meets expert standards (summative).

Worth evaluations focus on the stakeholders views to derive standards. To determine if a course under development is meeting local needs the values and needs of the local situation are employed as the basis for deriving standards. For a completed course, a needs assessment of the local situation serves as the basis for standards derivation. In both of these cases this information is determined in consultation with stakeholders in order to develop criteria and standards for evaluation.

In practical applications, such as TQM and other quality initiatives, similar sources are employed to derive standards, however they are often based on multiple sources rather than one or two specifically. Sources used include customer or stakeholder needs, organizational goals, professional/expert opinion, and benchmarking with other comparable institutions.

Kennedy (1994) provides insight into the derivation of standards and criteria for distance education course evaluation. The approach described consists of three steps

which are carried out by the evaluation team.

Step one deals with "audience concerns and issues" and entails the identification of "all possible audiences and stakeholders" in order to determine the information needs or issues that they consider pertinent to the evaluation (Kennedy, 1994, p.3). The second step consists of the analysis of all documentation, from proposals, to budgetary information, to course materials. Kennedy (1994) suggests this analysis allows the team to "establish the perceived goals and objectives of the program, the resource allocations, the program design and plans for implementation" (p.4). In the final step literature on distance education course design and subject matter specific information is reviewed.

Based on the information from all three sources standards and criteria for the evaluation are derived. Kennedy (1994) suggests that these standards and criteria should be ratified by at least the clients and if possible by all audience groups. This approach employs sources for standards derivation that have been described both theoretically and practically and lends support to the notion that in practical models of standards derivation, multiple sources may used.

Information for standards derivation can be gathered

through consultation with stakeholders/customers or through interviews, or questionnaires. Standards, objectives or primary indicators generally describe an ideal state of an aspect of a program, while criteria for success and secondary indicators refer to more precise definitions which can be measured to determine if a standard is being met.

Implications for This Study

Distance education has been firmly established as a means through which students can pursue post-secondary education. Autonomous distance education institutions have emerged worldwide. Considerable efforts have been made to evaluate distance education (Holmberg, 1986), however, difficulties have been identified (Evans, 1985). In many distance education evaluations there has been a lack of attention to the derivation of standards upon which judgments about the effectiveness of the program are based (McAnany, 1982; Keegan, 1986).

The use of standards in evaluation is identified as a pertinent issue by examining the literature in evaluation and TQM (Fink and Kosekoff, 1980; Fadale and Winter, 1985; Consortium for Policy Research in Education, 1993). Evaluation theories have addressed the importance of standards when judging program effectiveness (Stake, 1975;

Guba and Lincoln, 1981; Passow, 1987), however, few evaluation models specify how to derive standards.

The lack of linkage between evaluation theory and practice has also been identified by several authors (Lipsey and Pollard, 1989; Lewey, 1983). Provus (1969) states that "despite the title of a new educational periodical, Theory into Practice, there appears to be very little linkage between program evaluation conducted in public schools and the kind of theory described in universities" (p.244). Although, Provus was referring to schools over twenty years ago, the issue still remains the same. Evaluations are conducted in higher education institutions and value judgments are made, but on what basis? And how does actual evaluation practice in institutions relate to evaluation theory and models? Worthen and Saunders (1973) proposed that further research was needed in evaluation, and advocated examining other professions and disciplines. It is proposed that by examining the evaluation process as practiced at an autonomous distance education institution, we may be able to learn more about the derivation and application of standards, and their relationship to evaluation theory and quality practices described in the literature.

CHAPTER THREE

Methodology

Approach

A case study approach was employed in this study. Merriam (1988) describes case study as based on the naturalistic or qualitative paradigm. She states that "a qualitative case study is an intensive logistic, description and analysis of a single phenomenon or social unit. Case studies are particularistic; they are descriptive; they are heuristic - that is, they offer insights into the phenomenon under study" (p.21).

The case study approach was used to investigate the evaluation methodology employed at a distance education institution. The case, then, was the university. Specifically, it addressed the derivation and application of standards within the evaluation process. This approach is deemed appropriate, given the lack of research on evaluation standards in autonomous distance education, and for that matter in general.

Specifically, an interpretative case study was employed. This is described by Merriam (1988) as "descriptive data used to develop conceptual categories or to illustrate, support or challenge theoretical assumptions

held prior to the data gatherings" (p. 28). Merriam (1988) expands by stating that if there is a lack of related theory prior to data gathering that the researcher will "gather as much information as possible with the intent of interpreting or theorizing about the phenomenon" (p. 28). This process was employed in this case study.

Data Collection Methods

Initial contact with the institution consisted of meeting a contact member, who has responsibility for course evaluation, and conducting an exploratory interview. Merriam (1988) states that this type of interview is employed in order to "learn enough about a situation to formulate questions for subsequent interviews" (p. 74.). The data gleaned from the exploratory interview was checked against the general interview guide previously developed by the researcher to indicate areas of omission or redundancy.

Documents of past and current evaluation activity were reviewed to search out evaluation standards, both explicit and implicit. Types of evaluation documents were categorized so that differentiation could be made between ongoing monitoring activity and evaluation activity, both formative and summative.

Participants in this case study included the following

groups, all of whom are professional employees of the distance education institution:

a. program administrators/coordinators
b. instructional designers and program development team members
c. evaluation division committee members/employees
d. course/program instructors

Patton (1990) suggests that qualitative research focuses in-depth on small samples that are actually selected purposefully. In keeping with this notion, participants in this study were selected with the assistance of key informants in order to interview those individuals that were rich in information. Each participant was asked to suggest individuals that the researcher should interview regarding the topic of study. When interviews could not be accommodated with individuals due to other commitments, those individuals were then sent a questionnaire.

Data gathered from participant groups was through one or more interviews. All interviews ranged from semi-structured to unstructured as the case study progressed. In preparation for these interviews, an indicator interview guide was developed (See Appendix A). The guide was not

designed to be used in total, or in the sequence indicated. Rather, suitable questions from the guide were used on an individual basis, as the need for information was indicated.

A second instrument was developed based on the interview guide (See Appendix A). It is a questionnaire which was employed to collect data from individuals who were unavailable for interviews but may have had valuable insight into program evaluation at the institution. The questionnaire was piloted on three individuals, consisting of subject matter experts in both evaluation and distance education, prior to use. Recommended revisions were incorporated into the document.

The focus of the interview guide and questionnaire was:

a. Whether a specific evaluation approach or approaches were applied consistently to distance education courses and programs.
b. Whether specific standards were used as a basis for judgments, on criteria applied in the evaluation approach.
c. On what basis the standards employed had been derived.

Time Line

October, 1994 - March 1995

Completion of literature review and initial contact with the distance education institution.

April - September, 1995

Completion of data collection.

October - June 1996

Analysis of data and completion of formal thesis report.

Data Collection

A total of eight interviews were conducted. One interview consisted of an exploration of evaluation practices in general at the institution. Seven interviews entailed examining evaluation criteria specifically in terms of whether they were employed for evaluation, what type of evaluation they were employed for, how they were measured, and where this measure was derived from. This is in accordance with the interview guide at Appendix A. Responses from the interviews were collated and categorized for each criterion.

During the process of data collection it became evident that participants did not fit neatly into one of the four groups of professional employees. Two participants were administrators but also developed programs and instructed. Three participants were primarily instructors but were also involved in course/program development and evaluation. One participant was an administrator and was informally involved in the evaluation of support services. A final participant was primarily involved in course and program evaluation.

Ten questionnaires were distributed to course/program administrators/instructors with whom an interview was not conducted. Participants were not asked to identify themselves on questionnaire to ensure confidentiality. Six questionnaires were returned; however, only four were completed.

In addition, several informal interviews were conducted via telephone to explore or follow-up on information that had been collected.

Data Analysis Techniques

Data collected is analyzed on two levels, as described by Merriam (1988). Preliminary data analysis is semantic content analysis in accordance with Krippendorff (1980). The result of this level of analysis is a narrative report.

A second level of analysis involves the development of conceptual categories or themes in order to interpret or theorize about the phenomenon with reference to evaluation literature.

Chapter Four

Report and Analysis of Results

Introduction

The primary objective of this study was to explore the evaluation practices at a distance education institution. More specifically, the study examined what type of evaluation is taking place, what criteria are employed to measure standards, how these criteria are measured and from what sources are they derived.

Organization of the Findings

The findings are reported chronologically as they were collected. Initial data collection consisted of the analysis of documentation relevant to program and course evaluation at the distance education institution and is therefore presented first.

The majority of the remaining data was collected by employing two instruments: an interview guide and a questionnaire. Each instrument permitted both open-ended and closed responses from participants. Both instruments were designed to examine a series of evaluation criteria clustered into the following groupings: enrollment/

attendance criteria, cost criteria, learning criteria, support criteria and course design criteria. Due to the parallels in the design of the questionnaire and interview guide, data from each are presented together. Additional data incorporated into this section were also collected through informal telephone discussion.

During the progression of the study, it became apparent that some data referred specifically to comprehensive program evaluation while other data referred to course evaluation, with only slight overlaps to programs. Therefore, data in each criteria grouping is presented first under the heading pertaining to course evaluation, and secondly under the heading pertaining only to comprehensive program evaluation.

A final section on Performance Indicators addresses data that were collected through informal interviews, telephone discussion and the collection and analysis of a task force report.

Documentation Analysis

Documents from a variety of sources were obtained and analyzed. Data from each document or document grouping are presented under its respective heading.

Evaluation Reports

Five evaluation reports documenting comprehensive program evaluations, that had been conducted since 1987, were analyzed. Although the focus of the evaluations ranged from nursing education to student monitoring systems, the evaluation approach employed in each case was stakeholder based. That is, evaluation standards and criteria were developed by the evaluation team and stakeholders.

The methodology for data collection employed instruments such as questionnaires, interviews, and existing institutional data. As the development of standards and criteria were a function of the focus for the evaluation and the perspective of stakeholder input, there was variation among the instruments. For example, some of the standards and criteria employed include program use, program effectiveness and efficiency, program administration, perceived program success, registration, course availability and selection, academic advising, facilities, study time, assessment, course difficulty, external factors, self-evaluation, student services, course material, and delivery as well as whether the program was meeting identified needs.

Although each identified problems, the precise measure or indicator of success, which was deemed as acceptable or unacceptable for each criterion or standard was not

precisely indicated. Standards had been derived, but criteria were, for the most, part non-existent, or at least non-specific.

Evaluation Questionnaires

A sample of six student evaluation questionnaires from different distance education courses were examined. Some student questionnaires were designed to be completed on a unit by unit basis while others included the entire course but may refer to specific units or sections. Analysis of these questionnaires indicated that a variety of student satisfaction criteria are employed to measure the effectiveness of various course components. These are compiled under the relevant headings in Table 1.

Table 1. Student Satisfaction Criteria Employed in the Measurement of Course Components.

Course Components	Criteria
Course Materials: Study Guide	Utility in assisting comprehension Utility to indicate important information Readability Clarity of directions Level of interest Appropriate amount of information Effectiveness of format Areas of particular utility Areas of particular interest Areas requiring elaboration Areas of redundancy
Course Materials: Assigned Readings and Text	Readability Level of difficulty Adequate information Utility of illustrations Time spent reading Appropriateness of amount of reading
Assigned Work: Questions	Number completed Areas in which more questions are required Utility of questions Satisfaction with solutions provided to questions Inappropriate questions

Table 1. (Continued)

Course Components	Criteria
Assigned Work: Unit tests and Exercises	Usefulness Relevance Difficulty Comprehensiveness Interest Turn-around-time in receiving corrected test Usefulness of marker comments Satisfaction with Grading Preference for telephone or written marking Preference for telephone or written tests
Support: Tutorial Assistance	Number of contact times with tutor Satisfaction with the amount of contact Utility of tutor assistance Tutor knowledge Tutor communication skills
Overall	Satisfaction with the course Suggestions for improvement Area of the course most liked Area of the course least liked Plans to take future courses

The criteria employed in student questionnaires was measured using rating scales which varied depending on the course in review. A three point rating scale was most common. Examples of the descriptors employed for each point or level are provided in Table 2.

Table 2. Descriptors employed for three point student rating scales

Level 1	Level 2	Level 3
substantially	somewhat	minimally
very	so-so	not
generally	relatively	not at all
regularly	sometimes	not at all

In addition, questions were included which required a yes, somewhat and no scale and ratings from very poor to excellent. Many options for elaboration and open ended questions were incorporated into questionnaires as well.

Course Evaluation

Attendance-Enrollment Factors

Standard 1. Attendance and enrollment in courses are stable.

Criterion 1: Course Completion. All respondents indicated that this criterion was employed when conducting evaluations. One individual indicated that this criterion was examined, but at another level within the institution. In terms of what constituted success or an acceptable level for the completion criterion a range of percentages, from forty to ninety-five percent at the lowest and highest points, were provided. No less than a seventy-five percent course completion rate was cited as a successful course

completion rate by two individuals, however, for one person this percentage was based on calculations after the point where students have withdrawn. For the other individual the seventy-five percent applied only to students in the senior, higher level courses. For lower level or junior courses a course completion rate of sixty percent was deemed as acceptable. One other respondent indicated that a course completion rate of seventy percent was acceptable. Another indicated a course completion rate of forty to sixty percent was evident in a particular field and that due to the nature of the field, this course completion rate exists and is acceptable. It is important to note here that in the discussion of measuring course completion rates to determine success that all percentages discussed above are based on the number of people who complete a course in comparison to the number of people who initially register. The majority of individuals indicated that the measurements they provided were based, at least in part, on experience with past courses and data. Factors which came into play with past experience included characteristics of the field of study and consistency in course completion rates over courses and over time. Others indicated that the measurements were derived from a comparison to course completion rates throughout the university or to other universities who

offer similar courses. A final source provided for the derivation of this measurement was the use of research and stakeholder input in this area.

Overall, the levels required for an acceptable course completion rate were consistently high. The measurement of the course completion criterion, in general, appears to be affected by both the nature of the discipline and the level of studies being examined.

Criterion 2: Program Graduation. This criterion was unanimously described as very difficult to examine and interpret because many students take several courses through distance education in order to complete program requirements at another university where they do eventually graduate. Another compounding difficulty with this criterion was attributed to the different types of students which make up the population, such as those obtaining credit to learn specific information, those upgrading academic achievements and those returning to study after many years out of school. Due to the accessibility and flexibility of distance education courses many types of students exist take courses in addition to those who actually complete a degree or diploma program.

Despite these problems, some suggested that this

criterion can be examined by considering both the program graduation numbers and the course completion rate. Others suggested that this was an important criterion, but was examined at other levels within the institution. In any case the over-riding factor, in the analysis of this criterion, was that many students successfully completed the courses for which they registered but might not graduate from the related program.

The derivation of sources to measure program graduation rates included the use of existing data and university measures at the institutional level. Measures of institutional or comprehensive program evaluation are addressed later under that heading.

Criterion 3: Students Taking Future Courses. Similar to program graduation, the criterion of taking future courses was described as a poor indicator when evaluating courses. This is due in part to students taking courses to fulfill requirements at other universities, and as well because many individuals may take a group of courses in a particular area where they require some expertise or have some interest and once that need is fulfilled they no longer continue to access courses in a program.

Information gleaned from the university, stakeholders,

and research were considered when deriving a measure of success for this criterion. However, some respondents suggested that this criterion can be examined informally. One measure of this criterion consists of an expected range of twenty to twenty-five percent of students who return to take extra courses. This measure is based on university average of students returning as well as past experience with specific courses. Other measures employed to evaluate this criterion informally included tracking students and actually seeing what happens over time.

Criterion 4: Equal Access. The criterion of equal access applies consistently to courses and is a guiding principle for the university. The acceptable level of access for students is described unanimously as being at the 100% level and this measurement is derived from university mandate. At the course level, evaluation of this criterion is described as anecdotal or informal and may involve the analyzing the situation to ensure that access is provided to as many students as possible. Some respondents commented on the advent of technology and the impact on access. Although computers are entering into the realm of courseware access more and more, courses that do not require access to a computer continue to be available for those students needing

them. The only exception to this now is in an area of study in related technological disciplines which clearly require the use of that technology. However, pressures to employ technology to deliver courses are prevalent, which, it is suggested, may pose a threat to universal access in the future.

Other respondents indicated that 100% access is available, however certain hidden requirements existed. For example, a certain level of reading skills and even some writing skills, in reality, are required to succeed. Again, these are factors which may vary among disciplines.

Cost Factors

Standard 2. Cost per course and per student is feasible. Cost criteria included both cost per student and cost per course, however, through data collection it became evident that cost was examined primarily in relation to courses.

Criterion 1: Cost Per Course. The majority of respondents indicated that cost is examined and includes a variety of factors. Marking is one factor and there is a limit on how much money can be spent on tutor marking. One estimate provided indicated that $133.00 per course can be

allocated to marking and this plays a role in course design. Others addressed this factor in terms of the number of hours which could be allocated for tutor marking per course. Both criteria are set by the university finance department and are based on budget considerations.

Cost is also examined in course development in terms of textbooks, student manuals, study guides, copy write materials and must be within a given budget. An estimate of approximately $150.00 per course was indicated as the cost required to cover course materials. This criterion is also set from the finance department based on budget. This information must be calculated for each course. Courses will not be approved for development if the costs are too high.

Information based on the cost of course materials and tutor marking can impact on decisions about keeping courses when enrollment is low and costs are high. One of the reasons for this emphasis on the cost of textbooks and materials in the course development and evaluation process is because all required materials are included in the course costs. Unlike traditional institutions and universities, the students do not pay for texts and materials. The fact that all materials are included in course fees appears to be another reason that cost per student is examined primarily

at the institutional level.

Learning Factors

Standard 3: Courses result in desired learning outcomes

Criterion 1: Geared To The Learners' Level. This criterion is examined but is also described as very difficult to do at a distance. The Fleishe Readability Guide at a grade 12 or 13 level is employed by one respondent. This criterion is described as particularly important for junior courses which are expected to be at a level that requires no prior knowledge. Courses are developed and examined to ensure that this is the case. In addition, extra steps are taken to support the no prior knowledge measure such as including extra notes, explanations and references with the course materials. This entry level is derived based on the level of students in courses over time as well as student and tutor feedback.

Overall, this criterion is described as being examined heavily in course development. In the case of courses developed with other agencies as well as internally, input from steering committees is used to assess the learner level. In evaluation this criterion is evaluated based on feedback by both employers and students. It also appears

the measurement and learner level employed is influenced by both the level of study and the discipline or course in question.

In the past, there were other methods of examining this criterion. This included the use of instructional designers who read materials as persons not familiar with the material to see if it was at the right level. However, this no longer the case and materials are aimed at the median reading ability of students, and are developed and evaluated using the processes already described.

Criterion 2: Quantity Learned (tests and exams).

Quantity learned, in terms of tests and exams, is examined primarily on an informal basis. While there is a preference by some not to have tests and exams, more multiple choice tests and exams are in use. This is due to constraints on the time allocated for tutors to correct tests and financial limitations on how much money can be allocated for tutor services per course.

Difficulty also lies in determining what tests and exams really indicate. One method described as useful includes the comparison of a student's exam scores to assignments to see if there is a match. Others suggest a comparison of test and exam scores to previous grades in the

course. In addition comparison of scores among tutors is also examined to ensure some type of standardization in marking.

Exams, in general, are based on the objectives set and are not designed to contain unexpected questions. In distance education students are graded against specific criteria, as opposed to testing with the use of norm referenced procedures. Therefore, it is suggested that one can ascertain if 90% of the class is failing there is something wrong, whereas if 90% of the class is at the A level things are going very well. However, beyond this estimate, it is difficult to pinpoint a specific level required to satisfy this criterion, due to individual differences. Another problem with pin-pointing a level for this criterion is the observation that in many cases those students who do experience difficulty tend to drop out, and therefore course grades are typically high.

Others provided a range of levels required for the quantity of learning criterion. These are often used to see if scores are either too high or too low. A class average of 75% was described as the desirable level, in terms of measuring success by one respondent. This level is based on past experience with courses as well as comparative to other universities where the average is often 65%. Another

indicated that the required level consisted of a class average of 80% and this was based primarily on past experience with courses. A class average of 80% was cited again but in this case it referred specifically to scores on quizzes. For more in depth essay examinations an average of 60% was deemed as the acceptable level for this criterion. These levels were based primarily on comparisons to the same discipline in other universities.

Overall, the role of past experience and comparison to disciplines in other institutions impacted most on the derivation of levels for this criterion. While the examination of tests and exams themselves for level of difficulty was addressed, this was not employed by the majority in routine evaluation of courses.

Criterion 3: Quality of Learning. This criterion was described as being addressed when examining exams and assignments throughout the course on an informal basis. Questions such as adequacy of coverage of materials and ability of students to apply knowledge come into play. In addition, students' ability to solve problems is considered. Evaluation of this criterion was described as based on how well students could apply the knowledge or theory taught in a course.

Other indications of the quality of student learning in a course included how well students understood the material. This can be derived from the number of questions and the materials that students have difficulty with. The quality of assignments and essay work submitted is examined, and based on past experience and what is expected of the student, an estimate of the quality of learning and whether it is up to par can be determined.

Overall, in distance education there is a lot of one-on-one, individualized, instruction and therefore this type of evaluation is ongoing. If the majority of students are experiencing the same problem this would appear within one course run of six to eight months. As well, student and tutor feedback provide a great deal of information. A final source of information with which to evaluate this criterion is the numbers of students who are actually doing the work and succeeding at it. Based on all of this information, courses can be evaluated and modified.

Criterion 4: Student Acceptance To Graduate School.

This criterion is examined only on an informal basis or through observation in that professors are asked for references for admission to graduate school. Acceptance rates to graduate schools was estimated as 100% in one

discipline. However others suggested that they found this difficult to measure because so much depended upon which school or program that the student had chosen for graduate studies.

Criterion 5: Student Employment After Graduation. This criterion was only examined through observation for most respondents and not examined at this level by others. It was described as difficult in all cases, as so many students completing distance education courses are actually graduating from other universities. However, some programs which are in place in conjunction with external agencies do receive feedback on the employment of students who have graduated from their employers. But, this is not the norm for courses.

Criterion 6: Development of Individual Learning Skills. This criterion is examined when evaluating a course on an on-going basis by examining assignments and course work. It is addressed in course development and through ongoing monitoring by examining whether there adequate explanation and support for students in junior courses and less in those more senior, where it is expected that independent learning skills will have begun to develop. In more skill specific

disciplines, basic and complex skills are identified and the evaluation of individual learning skills can be measured by examining student attainment of these skills.

The success of this criterion is measured by the majority through the examination of exams and tests to verify if they are proceeding according to the normal pace. The normal pace is based on past experience with a given course. Others evaluate feedback from tutors, assignments and questions. Again, variations from the norm are examined and this norm is based on patterns which have developed in courses over time.

Criterion 7: Self-instructional Materials. This criterion is examined during course development and on an on-going basis. Any recurring difficulties are noted and acted upon. The standard, cited by respondents, for this criterion is a 100% level of self-instruction in materials. This is derived from the fact that courses are being completed at a distance and therefore this level is needed.

Materials are expected to be completely self-instructional. However, it is important to note that students do have access tutor support as well as help from the instructor. Courses are evaluated in course development and on an on-going basis to ensure all materials,

instructions are included. The effectiveness of the materials on an on-going basis is primarily judged based on both questions and phone calls to the tutor and professor. Questions are reviewed and difficulties are determined which can then be addressed.

Support Factors

Standard 4: Course support is adequate.

Criterion 1: Tutorial Support. Tutorial support was described as being examined when evaluating courses by all respondents. Students are assigned a tutor, and in terms of a level of support, they must be available to students at given times three days a week. This appears to be derived from university policy. Others describe a level of 100% in terms of accessibility by phone in the times provided. This is deemed as critical due to the nature of distance education. Feedback from students about tutorial support and availability are used as the primary measure for this criterion and complaints are examined on a recurring basis. The only difficulty cited in measuring support in this way is if students are receiving a busy dial tone, the tutor is not available but may be on the phone with another student.

Criterion 2: Tutorial Feedback and Timing. The acceptable measure for turn-around time for the return of assignments to students consists of work being received, corrected, and put back in the mail within a maximum of five working days. This level was described unanimously by respondents and is derived from university policy. Measurement of this criterion was deemed as feasible but somewhat tedious. Procedures included both verification through the registry for exams and through recording arrival dates of work submitted and postmark dates on outgoing mail.

Criterion 3: Organizational Support. Respondents indicated that students have access to a toll free line to contact both support staff and professors for any problem solving required. In addition, the library can be contacted, in a variety of ways, and support for students in terms of searching for and sending out materials is at the 100% level. This high level is required because of the nature of distance education, and because for many students this is the only library that they have access to. This level appears to be derived based on the university mandate of access for as many students as possible.

Another source of support is through email. In terms of measuring the support criterion, information is usually

obtained informally from students or tutors about difficulties experienced or complaints with a service. These are examined on a per course basis and changes are made accordingly.

Student Satisfaction Factors

Standard 5. Students are satisfied with the learning experience. Student satisfaction criteria are addressed continuously through the use of course evaluation questionnaires. Respondents indicated that the acceptable level of student satisfaction on these criteria consisted of receiving positive responses ranging from eighty to ninety-five percent of students. The exact measure varies depending on the field in question but is overall quite high. The measure and level required are based primarily on past experience with courses, however the use of literature and research findings were also addressed. While some variation from field to field exists, overall the student satisfaction criteria employed were similar in nature even if the terminology was not identical. A difficulty mentioned with the use of this data was that it only included students who completed courses. In addition, there is also the difficulty of how quickly this type of information can be responded to. This issue is addressed

further under course material - updatedness.

Criterion 1: Course Material - Essential. This criterion is examined in terms of the course objectives and what students are required to know based on them. Furthermore, this criterion is described as being heavily weighted in the course development process.

Criterion 2: Course Material - Interesting. Interest was also described as being an important criterion. However, the difficulty of ensuring that material was interesting to everyone, dependent on the field of study, was also evident. Some attempt to better address this criterion was made by trying to allow choice in part of the selection of materials so that students could choose that which was more interesting to them. Based on this, one can examine what students are selecting over time to determine what is more or less interesting.

Criterion 3: Course Material - Updated. This criterion was described as being examined on a one course per year basis due to workload and financial constraints. How often a course is evaluated varies, based on the number of courses a professor is carrying. For example, with a five course

workload, a course is examined every five years. Measurement of this criterion is also based on developments in the field. In an event where major changes were taking place in the field, it is expected that this would result in course evaluation and revision.

Course materials are described as being examined every four to five years by some and the requirement for implementing changes is based upon consistently arising difficulties. Others suggest every one to two years for minor revisions and seven to eight years for major revisions. Revisions can arise from changes to textbooks, developments in the field, or simply that a text is no longer being published. Overall, the data indicates that although any negative input from students is monitored, it may be difficult to respond to student satisfaction input quickly.

Criterion 4: Course Material - Presentation. This criterion is included in many course evaluation questionnaires but refers primarily to those items included as course study materials.

Criterion 5: Adequate Time To Complete Materials. This criterion is described as being monitored, however it is not

viewed as a problem. The expected completion time is set at six months to complete a course, and even then it is possible to get an extension. This has been derived based on the course loads and the special needs of distance education students who typically are leading very full lives and fitting their educational efforts into busy schedules. The opposing view of this criterion is the flexibility for students to finish very quickly, such as within two months, if they so desire. While this does not happen often the flexibility is there for students who are motivated to take advantage of the option.

Criterion 6: Comprehension. This criterion is examined on an ongoing basis in terms of student success on quizzes. However, in terms of student satisfaction major revisions are often also based on recurring student responses to questions on difficulty level.

Criterion 7: Appropriateness of Assigned Work. This criterion is addressed by students in terms of feedback about assignments during the course as well as on evaluation questionnaires. It is also examined in terms of measuring whether assigned work is a reflection of the objectives and is leading students to attain them. The objectives are

described as the yardstick against which assigned work is matched.

Criterion 8: Appropriateness of Learning Objectives. This criterion is addressed in evaluation questionnaires when students respond to whether they've used other materials and comment on learning materials in general. In addition, objectives are looked at, by some, in terms of whether they accurately reflect the course and difficulty level. Generally, lower level learning is required for junior courses and higher level for more senior courses. A great deal of this type of assessment is done when designing courses and it is expected that considerable comment would be required for this criterion to be examined beyond the course development process.

Criterion 9: Appropriateness of Grading. Feedback regarding the appropriateness of grading is gathered from questionnaires, tutor and instructor feedback. Although precisely defining an indicator of success for this criterion was described as difficult, two respondents did provide estimates. The first suggested that 90% of the students should be satisfied with the grading. Another suggested that if more than 15% of students comment

negatively on assignments or tests they should be examined as a critical part of the course.

Criterion 10: Appropriateness of Audio-Visual Support. This criterion is described as only being employed when it can enhance the learning experience and can then be evaluated for its contribution to a course. For example, audio-visual support has been employed in some courses to provide scenarios or simulations to students.

Course Planning Factors

Standard 6: Courses are comprehensively planned and based on established needs.

Criterion 1: Need for and utility of the course. The need for the course is described as being examined in terms of both developments in the field and needs identified by students. Included are students needing to complete program requirements, and areas where there is high interest in a course.

Generally, the criterion is described as difficult. Sometimes one can take a sample of students and ask them if they would be interested in a particular course, however this is not often done. Normally within the program a set

number of courses are required. Other universities offering the program are sometimes examined in terms of their courses and what is needed to get a benchmark. Needs assessment is informal.

Criterion 2: Economic Viability of the Course. The economic viability of a particular course is described as difficult to ascertain. Some respondents suggested numbers at other universities are examined. Overall, one must attempt to verify that 50 to 100 students would take a course within a given year in order to justify developing it. However some courses are needed for a particular program, despite numbers and economic viability.

Criterion 3: Validation of Course Materials Prior to Use. Field testing of some materials or courses is conducted but every course does not undergo this process as it would normally take a year to complete. Upon completion of field testing revisions are made, however, most often the field testing and revisions are completed only on a particular topic rather than the entire course. Field testing is described as having been carried out consistently at one time. In this scenario, an individual would go through materials unit by unit and try them out with

students, with revisions made as required. However, this is not often done now.

Summary

Overall, there is a great deal of course evaluation being conducted. Several respondents who had been at the distance education institution for many years did suggest that there was a greater focus on course evaluation in earlier years. At that time there were personnel who were devoted full time to field testing and other evaluation activities. The decline in evaluation activity is likely due to budget restraints and personnel reductions and reflects the current situation in many higher education institutions. In addition a small number of respondents expressed concerns about the difficulty of trying to implement timely changes based on the course evaluation feedback, noting that data collected and how it is used varies from one context to the next within the institution.

Comprehensive Program Evaluation

Evaluation Type and Derivation of Standards

Comprehensive program evaluation at the distance education institution was described as consisting of both

formative and summative evaluation. The derivation of standards, criteria and measurements are generally based upon the issues which concern the stakeholder groups who are a part of the evaluation process. Furthermore, the development of standards is described as the first order of business when conducting a comprehensive evaluation.

Comprehensive program evaluation, in general, is described as being initiated both internally or externally and can be based on a contract or in relation to a pilot course to assess its implementation. The following narrative discusses the criteria employed for comprehensive evaluation in relation to those in the interview guide and questionnaire, as well as additional criteria that were described by respondents. It is important to note that these criteria are not all inclusive. That is, others may be employed in comprehensive program evaluation depending on the situation and identified issues and concerns.

Attendance-Enrollment Factors

Course Completion. This criterion is examined in comprehensive evaluation and is described as being of interest to the institution and the government. It is measured using a comparative weighted measure of the

students in order to compare like with like. For example, students enrolled in similar courses are therefore compared to each other. Other measures described for this criterion included a comparison of the completion rate for a program to institutional averages as well as to resources expended.

Graduation Rates. Employing graduation rates in evaluation is described as difficult in distance education. However, two measures have been employed. The first is described as a cohort measure where the number of students in a program is examined in year one, year two and year three. This is examined in relation to an average which could be 100 students in year one, seventy in year two and fifty in year three. Based on a comparison a judgment can be made. This measure, although common, is described as not applying very well to distance education as students come in and out of programs continuously, creating a situation where there is no real dropout rate. The degree to which this impacts on distance education was evident in a recent study of two universities. The study indicated that approximately eleven hundred of students enrolled in these two universities graduated from programs using distance education credits from this distance education institution to complete program requirements.

A second measure is termed as full time equivalent and attempts to address the problem of measuring graduation rates in distance education where students may only take so many courses. It consists of determining the number of full time equivalent students at the institution and comparing this to the number of students graduating within a year.

Equal Access. This criterion is described as not being a major concern overall when conducting comprehensive evaluation because courses, programs and the institution as a whole, are designed to be open at a one hundred percent level. This criterion can become an issue when evaluating whether to use strategies such as seminar support. It can cost a lot and yet the only students who can use it are already relatively advantaged because they are located in major centers. Factors such as this are an issue because, although it is easier to enhance distance education in urban areas, this is not really the mandate of the institution.

Cost Factors. In terms of examining cost in program evaluation, the criterion employed primarily is the cost per successful course completion. It is described as a major measure due to the influx of people in and out of programs and the need to account for that factor. Further

measures of cost include a comparison of the cost per successful course completion to both the resources and support expended. Measures of cost can be made through comparison across courses, programs or to other institutions depending on the needs identified by the stakeholders and evaluator for a particular evaluation.

Learning Factors

Course Aimed At The Learners' Level. This criterion is described as being examined occasionally in program evaluation when it is warranted. Measurement is described as examining student progress and feedback in courses. Information to evaluate this criterion is drawn from instructor, university/stakeholder feedback, and perhaps even more heavily from the research literature.

Test and Exam Scores. The quantity of learning is examined in program evaluation and refers to criterion based testing which is employed for distance education at this institution. Evaluating this criterion involves examining pass rates, and this is an indicator that is continuously assessed. Generally, in distance education, students tend to drop out rather than to skip exams or to write when

inadequately prepared, as opposed to conventional universities. Therefore, the pass rate of those sitting exams is described as quite high at approximately 96%. The average grade is also high and is estimated at 75%, with little standard deviation from the norm. This is attributed to the fact that distance education evaluation is similar to mastery learning, in that students must complete all requirements to pass a course. Therefore this indicator is described as being poor, however it is monitored and it major deviations would be noted if they occurred.

Student Employment After Graduating. This criterion is described as being examined specifically from time to time. Measurement of this criterion consists primarily of a comparison of the employment of graduates of this institution to that of other institutions' graduates. The derivation of a specific comparative measure is based on the evaluation in question as well as research literature on this type of criterion.

Student Acceptance to Graduate School. Student acceptance to graduate school is examined occasionally and is compared with graduates from other institutions. However, there is difficulty with this criterion. It is

estimated that roughly twenty-five percent of graduates apply to graduate school. A description was provided of twenty-eight graduates from the distance education institution who had recently applied to graduate schools. Of the twenty-eight, twenty were accepted, two were waiting and six were not accepted but they were trying to gain entry in oversubscribed programs with high rejection rates. Therefore it is very difficult to indicate a level, across all programs, as to what is acceptable. There are so many factors that come into play, such as the program applied to and who happens to apply in a given year, that this criterion is difficult to specify in terms of a precise measure of success.

Development of Individual Learning Skills. As a specific criterion the development of individual learning skills is not really examined. More commonly, this type of issue may be examined in terms of English language proficiency and student study skills. Information on these factors is acquired through open ended questions in evaluations which focus on things such as difficulty. It is indicated that in their responses, some students will admit to requiring pacing as provided in a regular classroom rather than a self-paced environment. It is suggested that

if a high dropout rate exists for a course in a particular program, it is evidence that the method employed for instruction is inappropriate. As courses are highly print based, a high literacy rate is described as necessary.

Support Factors

Technical Access. Technical access to telephones, computers or televisions is described as being available to students, however most do not have any requirement to use these technologies to complete courses. Those that do require access to computers and telephones consist primarily of students studying computer science and distance education, because these students are pursuing studies in which technology is integrated within the discipline. The level required for access on the part of students is one hundred percent. This is described as not really an evaluation issue, as it is provided for in the university policy.

Organizational Support. Organizational support is examined in evaluations at this institution and is described as encompassing support components such the registry, computer help desk and tutor support. These factors are

rated by students as part of evaluation. However difficulties with these measure are addressed in terms of how one can interpret eighty percent of students providing positive responses about support and another twenty percent responding negatively. This difficulty of interpretation is attributed to the differential use of support by students. Some students use the support services infrequently but say it is very good, while others may have a tendency to give answers that they think the institution wants.

As the current measures do not indicate what problems exist very well, a new approach is described as being under development. This new approach consists of developing an Office of the Ombudsman to provide a forum for quick action when students have problems. It is expected that, through this means, the number of calls and the problems identified can be tracked and thus provide a more concrete performance measure to guide changes and improvements.

Tutorial Feedback and Timing. This criterion may be examined in terms of turn-around time. The measure of success for this criterion is five working days in and out, and is based on university policy. For evaluations it can be easily measured for exams, as tracking of exams is computerized. Measurement of this criterion is described as

much more difficult for assignments and feedback between tutors and students.

Student Satisfaction Factors

Measures of student satisfaction employed for evaluation are described as consisting primarily of student responses to questions about the academic or instructional quality of courses. This often entails having students respond using rating scales about various course components.

In comprehensive evaluation the criteria are described as stakeholder-based and are derived from all stakeholder groups. In relation to student satisfaction criteria, the difficulty of setting an absolute level is attributed to the fact that students do not operate within a closed environment. That is, students can have a variety of work and family variables impacting on their situation. Overall, when examining student satisfaction feedback on the criteria used for a particular evaluation, it is the extremes that are identified and analyzed to make judgments. A few of these criteria are addressed specifically.

Appropriateness of Learning Objectives & Grading. The development of learning objectives and their appropriateness is described as being examined extensively. However, this

occurs in phase three course development and is not really addressed further as a summative evaluation question. The appropriateness of grading is also addressed in course development and therefore has not been an issue for summative evaluation.

Appropriateness of Audio-visual Support. Audio-visual support is described as applying only to a minority of courses at this time. However, this type of support is being examined more extensively in language courses, and therefore may be examined as an evaluation criterion within that context.

Course Planning Factors

Course planning criteria, in general, are not examined because the institution uses a course planning development system. For each course in development a phase three report is written. This normally consists of a sample lesson, a course outline and a proposed student evaluation method. This report is circulated to the faculty, tutors and external academics, if required, for comment about the proposed course. Once circulation is complete, this report is returned to the Dean who will then approve the course or indicate what changes should take place. This formative

evaluation process has two purposes. It is intrinsic to the quality control of course planning, and also serves as method of communication to keep faculty informed of developments within the institution. A few evaluation criteria are addressed under this heading.

Need For The Course. The need for the course, as an evaluation criterion is described as not having been addressed. However, it is pointed out that this criterion should be looked at more extensively at the evaluation level. Currently, it is examined in the phase three course development phase and, it is suggested by one respondent, may consist of document collection, interviews and the presentation of arguments.

Economic Viability of the Course. This criterion, it is suggested, is examined and can consist of some type of cost measure. It is indicated that this criterion will be intensely examined in the future in terms of whether a course attracts a range from 100 to 1000 students, because of the realities of fiscal constraints. As an institution, it is possible to offer a finite number of courses, and it should be considered that these courses must eventually be revised, which incurs further costs. Revisions in the past

are described as occurring at one third of the courses per year, because of both manpower and budget constraints.

Cost of Materials and Development. It is suggested that the cost of materials and development is normally addressed in the phase three course development report. In this context it is examined in terms of the cost of student manuals, texts, and the overall development of the course. In terms of program evaluation the marketability of courses is the issue that receives attention and examination.

Validation of Course Materials. While the field testing of courses was common in the 1970s and 1980s it is described as occurring only occasionally now. When it does take place it normally consists having four or five students work through a new course. However, it is suggested that because validation is done so rarely, it would not be considered as a criterion in program evaluation.

Other Criteria

Other criteria were provided by respondents that have been utilized for evaluation purposes. Many are described as outputs which have been employed to evaluate distance education programs, and include criterion such as the cost

per successful course completion per graduate, as mentioned earlier.

Occupational development has been addressed by analyzing the jobs that students have obtained following graduation. Based on this information a comparative measure against other programs offering the same knowledge and skills is employed. It is pointed out that this can be difficult when someone is already working prior to course completion. As well, it is evident that there are not a significant number of distance education institutions with which to make these comparisons.

Another type of evaluation measure is conducted on a year to year basis. It is described as an internal historical comparison which provides an indication of how programs are doing.

Overall, there are different expectations for different programs and this usually impacts on evaluation and the criteria and standards employed. An example provided of one evaluation indicated that the stakeholders' primary concern was to see whether there was a rehabilitative effect resulting from the distance learning. This is not, it is ascertained, a distance education institutional indicator. However it was part of the program evaluation because of stakeholder input. Another example provided was of a

program where stakeholders wanted to see that the courses and materials were adapted specifically to their culture. Again, this is not necessarily a distance education evaluation criterion in general, but enters into evaluations through issues addressed by the stakeholders.

The mandate of the distance education institution is described as providing quality education to those who would not normally have it. However, in comprehensive program evaluation, the criteria are elicited from all stakeholder groups. Often they will have conflicting expectations, among themselves and with the institution. The process of setting measures is described as being somewhat complicated, as it is not easy to get everyone to agree on them. And in the final analysis, it is the evaluator or the evaluation team that must make a judgment call to determine standards and criteria or other factors that the stakeholders cannot agree upon or fail to specify.

Performance Indicators

The implementation of a performance indicator system within the distance education institution was addressed by a variety of respondents. Performance indicators are expected to affect evaluation by superseding comprehensive summative evaluation to some extent, as they are very specific. Also

the amount of resources required for conducting comprehensive evaluation is a factor in moving to a performance indicator system.

It is projected that the distance education institution will develop performance indicators of their own, which will be used to evaluate programs and the institution as a whole. This is in addition to those that are government-mandated. Comprehensive course evaluation, it is suggested, will probably only be done when demanded. With a solid system of indicators and sound management it is anticipated that comprehensive evaluation will probably not be needed on a routine basis.

Targets or measures that are set in this new system will depend on the indicators. Clearly the budget will provide one or more measures in the evaluation of programs. In addition, it is expected that the graduate full time equivalent (ie. the ratio of program graduates to full time equivalent enrolments in programs) will be used as a comparative target. For a new program, it is anticipated that interim and long term targets will be established so that one can measure to what extent people are leaning toward them and the degree of progression that is occurring. Overall, targets will be set by the governing council and the institution, since these are clearly the stakeholders.

However, it is expected that targets will also be affected by government input, as it is anticipated that they will use this information to cancel or advance programs.

Assessing the contribution of distance education in terms of access will still remain difficult. While students can be asked if there are barriers to conventional education that make access to distance education essential for them, this factor can vary from year to year. As a result, one cannot set a definitive target for this indicator. Although courses at a distance are available to students who do have barriers, this does not mean that they will enroll in them. Therefore, in a given year, it is conceivable that most students who take courses from the distance education institution are centrally located and can access conventional education. The distance education institution cannot control who decides to take advantage of the courses offered. Therefore, the requirement to provide access to courses through distance education has to be based on government mandate to provide quality education to those who would not normally have it.

Performance Indicator Report

The Report on Performance Indicators, that impacts specifically on this institution, outlines the development

of indicators in various areas such as access, student related indicators, community service, research, and fiscal management. Indicators have been developed by a task force consisting of stakeholders from all institutions who are affected. As the indicators for access already exist in the form of annual levels of full-time and part-time students, this area was not further addressed. It is essential to note that these indicators are under development and will be refined and improved based on pilot phase results. Therefore, this material provides a sense of the status of performance indicators under development, but is not in any way firmly adopted.

Student Related Indicators. Student related indicators which are under development for this distance education institution include student satisfaction and student persistence rates. Because of the nature of distance education many students do not graduate from programs. Therefore the indicator of student satisfaction may be examined through annual surveys of active students. In addition, because access is a critical mandate of the institution, surveys will also address the proportions of students working full-time or part-time and attempt to determine what, if any, barriers to conventional education

exist that make them choose distance education.

Student persistence rates for the distance education institution may be examined in terms of a ratio for student progress. The ratio compares the number of program graduates to the number of full time equivalent students (ie. those students identified as full time within the distance education context). Successful course completion rates are also considered as relevant to student progress and therefore may also be reported.

Research and Scholarly Performance Indicators. The indicators in this area apply to the distance education institution as well as to other institutions. Research may be examined in terms of publications and other creative work, publications without external sponsorship and federal research rates. The scope of "scholarly peer-reviewed publications and their creative and artistic equivalents include journal publications, books, book chapters, creative productions, contributions to conference proceedings, dissertations (terminal degree programs), patents and performances" (pg. 16.). A ratio may be employed to measure this indicator and consists of dividing the "total peer reviewed publications and other works by the total number of full time equivalent academic staff" annually (p. 16.). It

is suggested that this indicator be examined in combination with data on citation indices and the impact of publications. However this indicator is not considered as a reliable indicator for the humanities where monographs play a more critical role.

A similar ratio may be employed for publications without external sponsorship where the total number is divided by the number of full time equivalent staff annually. The scope of this indicator is identical to that in the first indicator except that in this case the work has not been sponsored externally. Any work funded externally, by federal and provincial grantors or supporters are not included.

Council success rates is another measure that may be applied to the distance education institution in this area. It is measured by dividing the total number of research grants awarded by the three national research councils in a period of three years by the average number of faculty members, in that time frame, who are eligible to apply. The councils include the Medical Research Council, the Natural Sciences and Engineering Research Council and the Social Sciences and Humanities Research Council.

A final indicator for Research and Scholarly Performance, specific to distance education, is under

development for inclusion in this area. The focus of this indicator will encompass the development of distance education systems, technologies and multi-media.

Cost Indicators. Cost is addressed in terms of cost per student, per year, per program, and cost per graduate, per program. Cost per student has two associated indicators. One examines the average cost on a per student, per year basis for each undergraduate degree program. The second examines the average annual cost on a per student basis for each master's and doctoral program in each faculty.

The cost per graduate, per program is examined in terms of the average total program cost per graduate, per program for each undergraduate, master's and doctoral program. How these indicators will apply, if differently, to the distance education institution based on the transfer of students in and out, is not addressed.

A sample student survey consists of both open and closed questions. Students are asked to respond to rating scales to evaluate services and facilities, learning experiences, faculty experiences, courses within and outside of their faculty and individual skills development.

Overall, this project is described as being at only the

starting point of determining how to measure performance. Furthermore, the use and interpretation of performance indicators is described as being complex and is therefore treated very carefully.

Analysis of Findings

Course Evaluation

Based on the interview and questionnaire data, it is evident that these criteria are applied primarily to course evaluation by the respondents interviewed at the instructor or departmental level. It is important to note that two respondents did indicate that some of the criteria are also applicable to program evaluation, but overall course evaluation was the norm. Course evaluation encompasses a range of evaluation approaches such as informal, ongoing and course-end evaluation, to formative and sometimes more formal summative course evaluation. The data clearly indicate that some form of course evaluation along this continuum is being carried out by all respondents. This is further supported by the research of Crawford and McGuire (1994) who document the collection of course evaluation data through student questionnaires, tutor records and student performance records. Crawford and McGuire (1994) note that analyzing and applying the data is more difficult, and discuss measures to address this problem.

While some slight differences exist from context to context, or field of specialization, which may impact on which criteria are used in a given situation, overall the

criteria employed are stable. For example, a criterion such as course completion is examined by the majority of respondents, while others such as student employment after graduation is only observed informally.

Table 3 displays the criteria grouping, its associated criteria and compiled data under the headings of: approach employed, measurement employed, and the derivation of the measurement. It is evident that all criteria are addressed with the exception of cost per student and the utility of the course. It would appear that one reason that cost per student is not examined is the fact that all costs such as books, computer disks, and the like are incorporated in the tuition fee of the course. The utility of the course was described as strongly associated with the need and economic viability criteria, and therefore is not included.

Several themes emerge from an analysis of the data. In terms of evaluation approach, it is evident that criteria fall loosely into two categories: those that are described as observed and those that are evaluated. Criteria that are observed or monitored, however, do not appear to have precise measurements associated with them at the course level. Even in the case where a measurement and an indicator of success is stipulated, evaluation of the criterion is described as tedious or difficult. Observed

criteria include students taking future courses for attendance/enrollment, student acceptance to graduate school and employment after graduation in the learning area, tutorial feedback and timing in the support area, and for student satisfaction course material presentation, appropriateness of audio-visual support and adequate time to complete materials. Informal monitoring or tracking occurs, as respondents do feel these criteria are still relevant

Of the remaining criteria, which fall into the evaluated category, three themes emerge. There are criteria that are associated with precise measurements and measures of success, those that are absolute or at the 100% level, and those that appear to be analytical or evaluated qualitatively. Those criteria that provide precise measures and indicators of success include: course completion, cost per course, economic viability and several student satisfaction criteria.

These student satisfaction criteria are evaluated based on the percentage of positive student responses on questionnaires. Respondents did provide precise measurements of success for student satisfaction criteria. It is also noteworthy that for the course completion criterion, many factors do impact on measurement. This was not the case for all criteria. For many criteria, stating a precise measure

or indicator of success was not deemed possible.

Table 3. Course and Departmental Evaluation Procedures

ATTENDANCE-ENROLLMENT CRITERIA	APPROACH EMPLOYED	MEASUREMENT EMPLOYED	DERIVATION OF MEASUREMENT
Course Completion	Evaluated	Percentage of course completions in comparison to the number of people who: initially register or remain after the withdrawal date. Range of success is 40% to 90%.	Course completion rates over time. Comparison to the institution course completion rates on average Comparison to other institutions with similar courses. Research. Stakeholder input. Past experience
Program Graduation	Evaluated	Comparison of course completion rates and program graduation rates. (Several factors impact)	Existing data and university measures.
Students Taking Future Courses	Observed	Percentage of students who return & take a course. Tracking students over time.	University average of students taking future courses. University records. Stakeholders Research
Equal Access	Evaluated	One hundred percent.	University policy.
Cost per Course	Evaluated	Hours/$ allocated for tutor marking, development or purchase of textbooks, student manuals, study guides, copy-rite.	Finance department budget

Table 3. Course and Departmental Evaluation Procedures (cont'd)

LEARNING CRITERIA	APPROACH EMPLOYED	MEASUREMENT EMPLOYED	DERIVATION OF MEASUREMENT
Geared to Learners Level	Evaluated	Grade 12 or 13 reading level No prior knowledge Feedback from students and employers.	Fleishe Readability Guide The level of students in courses over time Steering committee input
Quantity Learned (tests & exams)	Evaluated	Comparison of exam scores to assignments. Comparison of scores among tutors. Criterion referenced testing Percentages. Estimates of success as class average ranging from 60% to 80%.	Percentage is based on experience with courses over time. Comparison to the same discipline or overall average in other institutions.
Quality of Learning	Evaluated	Application of knowledge Comprehension Problem solving Student scores	The number of trainee questions Difficulty evident Assignments compared to past Student and tutor feedback.
Student Acceptance to Graduate School	Observed	By references and observation only	N / A
Student Employment After Graduation	Observed	By observation Occasionally some feedback from employers.	N / A
Development of Individual Learning Skills	Evaluated	Comparison of exams and tests to norm for past courses. Feedback from tutors, assignments and questions.	Patterns developed over courses over time.
Self-Instructional Materials	Evaluated	100%	Derived based on the need for complete self-instructional materials in distance learning.

Table 3. Course and Departmental Evaluation Procedures (cont'd)

STUDENT SATISFACTION CRITERIA	APPROACH EMPLOYED	MEASUREMENT EMPLOYED	DERIVATION OF MEASUREMENT
Overall	Evaluated or Observed	Positive feedback on criteria from a range of 80 to 95% of students.	Student response on questionnaires and other factors specific to a criterion.
Course Material - Essential	Evaluated	Examined in terms of course objectives.	Past experience with courses. Literature & research.
Course Material - Interesting	Evaluated	Student response is difficult to measure, so provide choice and track selections.	Courses over time.
Course Material - Up datedness	Evaluated	Negative input from students is monitored. Course revised every four to five years.	Developments in the field. Courses over time.
Course Material - Presentation	Observed	Student response	Courses over time.
Adequate Time to Complete Materials	Evaluated	6 months to complete a course plus an extension if required.	Based on course loads and needs of students. This is university policy.
Comprehension	Evaluated	Student performance and student response.	Student response and results on quizzes in courses over time.
Appropriateness of Assigned Work	Evaluated	Comparison to objectives. Student response during course and questionnaires.	Courses over time and student response over time.
Appropriateness of Grading	Evaluated	Negative student response is monitored	Responses on questionnaires over time and comments.
Appropriateness of Audio-Visual Support	Observed Evaluated more *when in use.	Only when it facilitates learning	N/A
Appropriateness of Learning Objectives	*Heavily evaluated in course development.	Difficulty level and whether material reflects objectives. Student response.	Student response over time and course objectives.

Table 3. Course and Departmental Evaluation Procedures (cont'd)

SUPPORT CRITERIA	APPROACH EMPLOYED	MEASUREMENT EMPLOYED	DERIVATION OF MEASUREMENT
Tutorial Support	Evaluated	100% access to tutors during time specified - Judged from complaints & tutor feedback.	University policy
Tutorial Feedback & Timing	Observed	Acceptable turn-around time is 5 working days to receive and send out work. Tedious to track but possible through post marks & registry.	University policy
Organizational Support	Evaluated	100% monitored through complaints and feedback.	University policy

COURSE PLANNING CRITERIA	APPROACH EMPLOYED	MEASUREMENT EMPLOYED	DERIVATION OF MEASUREMENT
Need for the Course	Evaluated	Requests for the course. Developments in the field. Course success at other institutions.	Program requirements. Benchmarking to other institutions.
Economic Viability	Evaluated	Numbers to take the course versus cost. One estimate of 50 to 100 students taking the course per year required.	Budgets Program requirements
Validation of course materials prior to use.	Evaluated	Field testing (rarely) External Evaluators Comparative to other Universities Advisory Committees	Stakeholder input Expert opinion Other institutions use.

The second category are those criteria which are evaluated and must be at the complete or one hundred percent level of success. These include equal access, self-instructional materials, tutorial and organizational support. These criteria have an essential relationship to the provision of learning within distance education. Due to the nature of distance education, the measures of success for these criteria is at one hundred percent.

Criteria which appear to require analytical or qualitative evaluation and do not provide precise indicators of success are program graduation rate, geared to the learners' level, quality of learning, development of individual learning skills, updatedness of course material, appropriateness of audio-visual support, need for the course and validation of course materials prior to use.

When examining the derivation of measurement for criteria, a variety of themes or categories can be drawn from the data. Measurements are derived based on:

a. Comparisons or benchmarking both internally and externally to similar disciplines at other institutions, averages at other institutions, averages within the institution, past exam scores within the institution, and past feedback.

b. Factors internal to the institution which include existing data, budget, program requirements, and institutional measures, records, averages, mandate, and policies.

c. Expertise such as literature, research, tools and expert opinion.

d. Stakeholder input which encompasses the personal insight and feedback of advisory committees, instructors, tutors, students, and employers (occasionally).

Overall, the derivation of a measurement of success for criteria was based on more than a singular factor and often crossed the boundaries of the four categories described above. The only exception are those criteria strongly associated with indicators of success at the 100% level. The measurement of those criteria is primarily derived from university mandate. However, although university mandate appears to be a single source for the derivation of measurement, it is highly probable that it is not. It is quite probable that measurements derived based on university mandate, which are clearly entrenched in the system, have been put in place based on a variety of factors. These would probably include distance education research,

literature, and stakeholder input at the institution. Therefore this derivation of the 100% level as a measure of success likely crosses the boundaries of the four categories, as does the derivation of measurement for the other criteria

From the evaluation literature review, it is evident that the derivation of standards is differentiated, by some authors, based on the purpose of an evaluation. The source of standards for merit evaluations, identified in Guba and Lincoln's (1981) naturalistic model, is some form of expertise or similar expert programs. That is, professional opinion and literature serve as the source for deriving standards for perfecting courses (formative) or determining if they meet expert standards (summative). Merit is distinguished from worth evaluations which focus on stakeholder views to derive standards. For a program under development (formative) the values and needs of the local situation are employed to determine if local needs are being met. For a summative evaluation of a completed program, a needs assessment of the local situation serves as the basis. In addition Guba and Lincoln's (1981) naturalistic model addresses program evaluation versus course evaluation.

A definition of what is meant by the local needs must be addressed prior to applying this theory to distance

education. Normally the local needs are easy to determine within the local boundaries. However, because distance education meets the needs of students from dispersed and sometimes isolated regions, the term local needs must be interpreted to include those students, personnel and stakeholders who are served by the institution, although they are not within a local boundary.

The notion of deriving standards for merit or worth evaluations can be applied to the derivation of standards data for course evaluation in this study, as is illustrated in Table 4. Categorization of the derivation of standards according to this framework indicates that the majority of the sources cited can be linked to either merit or worth type of evaluations as described by Guba and Lincoln(1981). For many of the criteria the source of standards derivation is linked to both merit and worth. While standards derived from university mandate or policy have been linked to worth evaluations only, it is quite plausible that the policies and mandates themselves may be based on literature and expert opinion, which then relates back to merit. It must be recognized that Guba and Lincoln (1981) address the derivation of standards for program evaluation and that the purpose of the evaluation is quite well defined. The data here is primarily for course evaluation and the evaluation

context is very broad. Perhaps if the purpose of the course evaluation was clearly delineated, the data on standards derivation may fall into the categories for merit and worth as described by Guba and Lincoln (1981). Only further specific study in this area could confirm this.

Table 4. Comparison of the Derivation of Course Evaluation Standards to Models & Theory

ATTENDANCE-ENROLLMENT CRITERIA	DERIVATION OF MEASURE	COURSE EVALUATION MEASURE	QUALITY MODELS	MERIT/ WORTH
Course Completion	Course completion rates over time. Comparison to the institution course completion rates on average Comparison to other institutions with similar courses. Research. Stakeholder input. Past experience	Percentage of course completions in comparison to the number of people who: initially register or remain after the withdrawal date. Range of success is 40% to 90%.	Baselines Benchmarks	Worth-Stakeholders and local needs. Merit-expert institutions, opinions, and literature.
Program Graduation	Existing data and university measures.	Comparison of course completion rates and program graduation rates. (Several factors impact)	Baselines	Worth
Students Taking Future Courses	University average of students taking future courses. University records. Stakeholders Research	Percentage of students who return & take a course. Tracking students over time.	Baselines Benchmarks	Worth & Merit
Equal Access	University mandate.	One hundred percent.		Worth-University Mandate (likely merit too)
Cost per Course	Finance Department Budget	Hours/$ allocated for tutor marking, development or purchase of textbooks, student manuals, study guides, copy-rite.		Worth

Table 4. Comparison of the Derivation of Course Evaluation Standards to Models & Theory (con'd)

LEARNING CRITERIA	DERIVATION OF MEASURE	COURSE EVALUATION MEASURE	QUALITY MODELS	MERIT/ WORTH
Geared to Learners Level	Fleishe Readability Guide The level of students in courses over time Steering Committee Input	Grade 12 or 13 reading level No prior knowledge Feedback from students / employers.	Benchmarks Baselines	Merit & Worth
Quantity Learned (tests & exams)	Percentage is based on experience with courses over time. Comparison to the same discipline or overall average in other institutions.	Comparison of exam scores to assignments. Comparison of scores among tutors. Criterion reference testing percentages	Baselines Benchmarks	Worth & Merit
Quality of Learning	The number of trainee questions Difficulty evident Assignments compared to past Student and tutor feedback.	Application of knowledge Comprehension Problem solving Student scores	Customer feedback Baselines	Worth
Student Acceptance to Graduate School	N / A	By references and observation only	N/A	N/A
Student Employment After Graduation	N/A	By observation Occasionally some employer feedback	N/A	N/A
Self-Instructional Materials	Patterns over courses over time.	Must be at 100% level	Baselines Benchmarks	Merit & Worth
Development of Individual Learning Skills	Derived based on the need for complete self-instructional materials in distance learning.	Comparison of performance to past. Feedback from tutor, assignments and questions.	Customer feedback Baselines	Worth

Table 4. Comparison of the Derivation of Course Evaluation Standards to Models & Theory (con'd)

STUDENT SATISFACTION CRITERIA	DERIVATION OF MEASURE	COURSE EVALUATION MEASURE	QUALITY MODELS	MERIT/WORTH
Overall	Student response on questionnaires and other factors specific to a criterion.	Positive feedback on criteria from a range of 80 to 95% of students.	Customer feedback Benchmarks	Worth & Merit
Course Material - Essential	Past experience with courses. Literature & research.	Examined in terms of course objectives.	Customer feedback Benchmarks	Worth & Merit
Course Material - Interesting	Courses over time.	Student response is difficult to measure, so provide choice and track selections.	Customer feedback Baselines	Worth
Course Material - Updatedness	Developments in the field. Courses over time.	Negative input from students is monitored. Course revised every four to five years.	Customer feedback Benchmarks	Worth & Merit
Course Material - Presentation	Courses over time.	Student response	Customer feedback	Worth
Adequate Time to Complete Materials	Based on course loads and needs of students. This is university policy.	6 months to complete a course plus an extension if required.	Customer needs	Worth
Comprehension	Student response and results on quizzes in courses over time.	Student performance and student response.	Baselines Customer feedback	Worth

Table 4. Comparison of the Derivation of Course Evaluation Standards to Models & Theory (con'd)

STUDENT SATISFACTION CRITERIA	DERIVATION OF MEASURE	COURSE EVALUATION MEASURE	QUALITY MODELS	MERIT/ WORTH
Appropriateness of Assigned Work	Courses over time and student response over time.	Comparison to objectives. Student response during course and questionnaires.	Customer feedback Baselines	Worth
Appropriateness of Grading	Responses on questionnaires over time and comments	Negative student response is monitored	Customer feedback	Worth
Appropriateness of Audio-Visual Support	N/A	Only when it facilitates learning	N/A	N/A
Appropriateness of Learning Objectives	Student response over time and course objectives	Difficulty level and whether material reflects objectives. Student response.	Customer feedback Baselines Benchmarks	Worth & Merit (in course development)

SUPPORT CRITERIA	DERIVATION OF MEASURE	COURSE EVALUATION MEASURE	QUALITY MODELS	MERIT/WORTH
Tutorial Support	University policy	100% access during time specified - Judged from tutor feedback, complaints.	Customer feedback	Worth - University Mandate (likely merit too)
Tutorial Feedback & Timing	University policy	5 working days to receive and send out track through post marks & registry.		Worth
Organizational Support	University policy	100% monitored through complaints and feedback.	Customer feedback	Worth but as is policy is likely also merit

Table 4. Comparison of the Derivation of Course Evaluation Standards to Models & Theory (con'd)

COURSE PLANNING CRITERIA	DERIVATION OF MEASURE	COURSE EVALUATION MEASURE	QUALITY MODEL MEASURES	MERIT/ WORTH
Need for the Course	Program requirements. Benchmarking to other institutions.	Requests for the course. Developments in the field. Course success at other institutions.	Customer needs Benchmarks	Merit & Worth
Economic Viability	Budgets Program requirements	Numbers to take the course versus cost. One estimate of 50 to 100 students taking the course per year required.	Baselines Needs	Worth
Validation of course materials prior to use.	Stakeholder input Expert opinion Other institutions use.	Field testing (rarely) External Evaluators Comparative to other Universities Advisory Committees	Customer feedback Benchmarks	Worth & Merit

In this study, the source of standards derivation is related to merit and worth, and often to both. Techniques employed for standards derivation can also be linked to the quality models described by Carpezio and Moorehouse (1993) and Lewis and Smith (1994).

Carpezio and Moorehouse (1993) describe the development of baselines as a source for standards of excellence within an institution. The use of baselines is evident in course evaluation within this institution, when measurements of a

particular criterion are examined over time and applied to a course to judge its success. This approach is employed for many criteria in each broad area of attendance/enrollment, learning, student satisfaction, support and course planning.

Benchmarking is a second measure described by Carpezio and Moorehouse (1993). It is employed when comparisons to other similar institutions or departments within the institutions are made for a particular criterion

The use of stakeholder or customer feedback in the development of standards or indicators is addressed by Lewis and Smith (1994). This is evidenced, at this institution, in the use student and tutor feedback in course evaluation.

A comparison of course evaluation at this institution to performance indicators was not addressed as performance indicators are generally described as used at the institutional or program level. While stakeholders do play a role in the derivation of course evaluation standards, the approach is clearly not solely stakeholder based. Indeed, in the final analysis, the decision as to what standards and criteria to employ is derived from many sources and may at time rest with the professor. The data suggest that multiple sources are employed for the derivation of standards and multiple measures, related to many models, are applied to distance education course evaluation at this

institution.

Comprehensive Program Evaluation

Comprehensive Program Evaluation is conducted at this institution in response to internal or external demands or requirements for the evaluation of a particular program. This evaluation may be both formative or summative in nature.

The focus of the comprehensive program evaluation is driven by the concerns and issues identified by the stakeholders who are part of the evaluation process. The derivation of standards, criteria and measurements are identified by Scriven (1991) as the arm of evaluation that "collects, clarifies and verifies relevant values and standards" (p.5). The focus on the stakeholder that is employed for comprehensive program evaluation relates strongly to the approach advocated by Guba and Lincoln (1981). They state that "responsive evaluation produces information that audiences want and need" (p.38). The use of a stakeholder approach to comprehensive program evaluation is further supported by the analysis of evaluation reports from this institution in the documentation analysis section.

The criteria employed are a function of stakeholder

issues and concerns. Although not all inclusive, those addressed as in use by respondents at this institution are identified as important for distance education by Paul (1990) and Gooler (1979). Furthermore, the criteria correspond to those at other distance education institutions for program evaluation (Bartels, 1991; Ganor, 1991; Koul, 1991; Woodley,1991).

The derivation of standards and criteria for comprehensive program evaluation indicates that there is a focus on precise measures and outputs versus inputs. For most criteria addressed, multiple measures have been applied to evaluate how well the standard is being met. Although a stakeholder approach is employed in program evaluation, the derivation of standards and criteria can be linked to both Guba and Lincoln's notion of merit and worth as well as other quality models. This relationship is illustrated in Table 5.

The derivation of measurement for program evaluation criteria at this institution can be related to Guba and Lincoln's (1981) notion of using stakeholder information needs and views to derive worth standards and using professional literature and expert opinion as the source of merit standards. Standards for particular criteria can be related to both merit and worth, and for many criteria

multiple measures relating to both source types are applied.

Table 5. Comprehensive Program Evaluation Procedures

ATTENDANCE-ENROLLMENT CRITERIA	PROGRAM EVALUATION MEASURE	QUALITY MODEL MEASURES	PERFORMANCE INDICATOR MEASURES	COMPARISON TO MERIT WORTH
Course Completion	Comparative weighted measure to students in general Comparison of the program completion rate to institutional averages Comparison of program completion rate to resources expended	Baselines		Worth-Stakeholder
Program Graduation Rate	Cohort measures Full time equivalents compered to students graduating per year	Baselines	Student Persistence Rates	Worth-Stakeholder
Equal Access	Only examined in terms of whether to employ things like seminar support (university mandate)			Worth-Stakeholder & Merit-literature, expert opinion
Cost per Course	Cost per successful course completion Comparison of cost per successful course completion to resources and support expended		Cost Measures	Worth-Stakeholder

Table 5. Comprehensive Program Evaluation Procedures (con'd)

LEARNING CRITERIA	PROGRAM EVALUATION MEASURE	QUALITY MODEL MEASURES	PERFORMANCE INDICATOR MEASURES	COMPARISON TO MERIT & WORTH
Geared to Learners Level	Student progress Feedback on courses (derived from literature and stakeholders)			Merit - Literature Worth - Stakeholder
Quantity & Quality Learned (tests & exams)	Test and exam scores: pass rates (of those sitting avg=96%) and average grade (avg=75%)	Baselines		Worth - Stakeholder
Student Acceptance to Graduate School	Compared to graduates from other institutions	Benchmarks		Worth - Stakeholder & Merit if other institution is considered an expert
Student Employment After Graduation	Comparison of graduates employment from other institutions	Benchmarks		Worth - Stakeholder & Merit if other institution is considered an expert

Table 5. Comprehensive Program Evaluation Procedures (con'd)

SUPPORT CRITERIA	PROGRAM EVALUATION MEASURE	QUALITY MODEL MEASURES	PERFORMANCE INDICATOR MEASURES	COMPARISON TO MERIT & WORTH
Organizational Support	Student ratings of: registry, computer desk, and tutor support New approach: Ombudsman / track number of calls and number of problems	Customer interview & survey feedback	Annual student surveys	Worth - Stakeholder
Tutorial Feedback & Timing	Five working days Exams tracked through computer registry (mandate)	Development of a standard of service that is monitored		Worth - Stakeholder & Merit - Literature and expert opinion

STUDENT SATISFACTION CRITERIA	PROGRAM EVALUATION MEASURE	QUALITY MODEL MEASURES	PERFORMANCE INDICATOR MEASURES	COMPARISON TO MERIT & WORTH
Varies from evaluation to evaluation	Student responses on rating scales Look at extremes	Customer interview & survey feedback	Annual student surveys	Worth - Stakeholder
Appropriateness of Grading	Examined extensively in course design	Customer interview & survey feedback	Annual student surveys	Worth - Stakeholder
Appropriateness of Audio-Visual Support	Only when used			
Appropriateness of Learning Objectives	Extensively in course design	Customer interview & survey feedback	Annual student surveys	Worth - Stakeholder

Table 5. Comprehensive Program Evaluation Procedures (con'd)

COURSE PLANNING CRITERIA	PROGRAM EVALUATION MEASURE	QUALITY MODEL MEASURES	PERFORMANCE INDICATOR MEASURES	COMPARISON TO MERIT & WORTH
Need for the Course	Has not been looked at extensively Document collection Interviews Presentation of arguments			
Economic Viability	Numbers to take the course versus cost. Marketability of the course			

OTHER	PROGRAM EVALUATION MEASURE	QUALITY MODEL MEASURES	PERFORMANCE INDICATOR MEASURES	COMPARISON TO MERIT & WORTH
Internal Historical Measure	Comparison of program data from year to year	Baselines	Data/records over time	Merit and Worth

The use of a stakeholder approach and multiple measures to derive standards for program evaluation at this institution can also be related to quality models. This relationship is illustrated in Table 5. Tannock (1992) expounds on the development of quality system standards and addresses not only academic but also support and organizational components which may be examined to assess a program. His views are particularly applicable to distance

education where support and organizational factors are considered important.

As well, Tannock (1992) advocates using inputs from other institutions, associations and business to help address these points. This notion of benchmarking is also advocated by Capezio and Moorehouse (1993). Although benchmarking is employed in comprehensive program evaluation at this institution, it is described as difficult because there is a lack of comparable distance education institutions. Therefore internal comparisons are often made, or baselines employed which look at current or past activity to derive a standard. The use of baselines is also advocated by Capezio and Moorehouse (1993) in total quality management models.

An approach which focuses on the internal and external customers is employed in both of these models and relates heavily to the overall stakeholder approach employed at this institution. This stakeholder or customer approach is also advocated by Lewis and Smith (1994) in applying a total quality management system to higher education. They contend that the measurement of output is critical. The use of measurements of output, advocated by Lewis and Smith (1994), is evident at this institution in terms of performance on exams, quizzes, assignments and pass rates.

A final comparison is made from standards derivation at this institution to the development of performance indicators and their standards or indicators of success. The development of some indicators do correspond or relate to those derived and applied in comprehensive program evaluation at this institution. These include the use of student persistence rates, cost factors, and student surveys of organizational, support and other student satisfaction criteria. Adams (1991) also advocates a stakeholder approach to the use of performance indicators to measure effectiveness and efficiency in higher education.

Summary

The use of various types of measurements for criteria indicates that many models or frameworks of evaluation impact on the comprehensive evaluation conducted at this institution. The impact of strategies other than those relating directly to Guba and Lincoln's elaboration of responsive evaluation is further evidenced in that, at this institution, it is the evaluator or the evaluation team that is the judge for the standards, criteria and measurements that are employed. This approach is more in keeping with the practical model outlined by Kennedy (1994) for the development of standards, criteria, and measurements for

comprehensive evaluation. Kennedy's (1994) approach employs multiple sources to derive standards for distance education program evaluation which include stakeholder issues, all documentation sources and related literature and research in the area. Although Kennedy's (1994) approach advocates the ratification of standards and criteria by all audience groups, it is the evaluator or evaluation team who ultimately decides which criteria and standards are employed.

Analysis of the findings suggests that while there is a strong relationship between practical and theoretical program evaluation, no single theory of evaluation is applied when conducting evaluation at this institution. This may be related to the many difficulties and variations from situation to situation which are encountered when applying evaluation theory in an actual program or institutional setting. While theoretical frameworks of evaluation, such as Guba and Lincoln's (1981), may have great impact on the evaluation approach employed, a single theory may not be able to address the needs of every evaluation situation. Therefore multiple sources, from both practical models and from theory, may be employed in the derivation and application of program evaluation standards. Analysis of the findings of this study suggests that this is

the case for distance education program evaluation at this institution.

The advent of the Performance Indicator System appears to be in response to the difficulty of conducting comprehensive program evaluation in a time of personnel and budget restraints, as well as, increased demands for accountability. Indicators addressed as being developed for use at this institution relate to those outlined by Adams (1993) as useful for program evaluation .

The application of performance indicators is partially mandated by government which forms part of the stakeholder group. Those indicators under development that are in addition to those required by government mandate, attempt to provide a more comprehensive indicator system for evaluation at this institution, and will attempt to meet the need for evaluation in a time of limited resources. This suggests that, in the future, performance indicators may act as a stimulus for comprehensive program evaluation when there is evidence of serious problems in a particular program. Whether this will result in the reduction of the need for, or amount of comprehensive program evaluation conducted overall, remains to be seen. As performance indicators are not the focus of this study, the impact of this system on this distance education institution is not addressed in

depth. Further analysis would require specific research in this area and is beyond the scope of this study.

Chapter Five

Conclusions and Recommendations

Summary

This study explored the program evaluation methods employed at one distance education institution, focusing in particular on the derivation and application of standards within the evaluation process. Using an interpretive case study approach, with the institution as the case, the researcher "gathered as much information as possible with the intent of interpreting or theorizing about the phenomenon" (Merriam, 1988, p. 128).

Through questionnaires, interviews and the review of institutional documents the researcher was able to ascertain what the institution used as standards and criteria in both course and program evaluation, and the derivation of these standards and criteria, as well as sources of data and measurement.

The data indicate that no one approach to the setting and use of standards and criteria is followed. Rather, input from stakeholders, the development of performance indicators by departments, university policy, use of baseline data, and comparative data from the literature and other institutions all contribute to the derivation of

standards and their specific criteria.

Conclusions

Course Evaluation

Analysis of the findings indicate that distance education course evaluation is conducted routinely at this institution. This finding is supported by the documentation analysis of student questionnaires. Criteria which are observed are not associated with precise measurements, while those described as evaluated fall into three categories. They include criteria that are associated with precise measurements, those that have a minimum standard of one hundred percent, and those that require analytical or qualitative assessment.

The derivation of standards and criteria for distance education course evaluation is based on comparisons or benchmarking, factors such as data internal to the university, sources of expertise and stakeholder input. Overall, the derivation of standards of success is based on multiple sources which cross the boundaries of the four areas described above.

The derivation of standards for distance education course evaluation can be linked to the sources for merit and worth program evaluation described by Guba and Lincoln

(1981). Often criteria relate to both merit and worth sources. The use of sources such as baselines and benchmarking to derive standards is evident and relates to quality models described by Capezio and Moorehouse (1993). As well, the use of student and tutor feedback relates to the use of customer feedback in the development of standards described by Lewis and Smith (1994).

This suggests that there is a relationship between the underpinnings of the standard setting process outlined in theoretical and practical models. Furthermore, the use of multiple sources for standards derivation and multiple measures for criteria is evident in distance education course evaluation at this institution.

Comprehensive Program Evaluation

Analysis of the findings suggest that the derivation and application of standards for distance education program evaluation at this institution is primarily stakeholder based. This is further supported by the documentation analysis of evaluation reports where stakeholder issues and concerns are addressed. Furthermore, the criteria described, although not all-inclusive, are identified as important for distance education, and in use at other distance education institutions, by several authors.

The stakeholder approach to the derivation of standards at this institution relates strongly to Guba and Lincoln's (1981) model of evaluation. Standards for criteria can be related to the sources identified for merit and worth evaluations described by Guba and Lincoln (1981). The stakeholder approach and sources in use such as benchmarking, baselines and customer/stakeholder surveys at this institution can also be related to quality models in the literature. This further supports the notion that a relationship exists between the basic underpinnings of the standard setting process across the theory and quality models described in the literature.

While distance education program evaluation at this institution is strongly linked to theory (Guba and Lincoln, 1981),the impact of strategies from other models is evident. This suggests that multiple sources, from both theoretical and practical models are employed to derive and apply distance education program evaluation standards. Furthermore, despite the stakeholder approach, it is the evaluator or evaluation team who ultimately decides which standards and criteria will be applied for evaluation at this institution. This relates strongly to the practical model of standards derivation for distance education program evaluation advocated by Kennedy (1994). Although there is a

significant relationship between theory and program evaluation in practice at this institution, other sources are drawn upon in the derivation and application of standards to meet the requirements of a given evaluation context.

Performance Indicators

A final comparison indicates that the derivation of selected performance indicators also correspond to those employed for program evaluation at this institution. However, the implementation of the performance indicator system is still under development. Therefore the impact of performance indicators on distance education program evaluation can only be determined through future research.

Recommendations

This study has provided baseline data on the derivation and application of course and program evaluation standards employed at one distance education institution. Furthermore, the comprehensive program evaluation conducted relates heavily to the evaluation theory described by Guba and Lincoln (1981) and is also linked to standards derivation in quality assurance and practical evaluation models. This suggests that both theoretical and practical

models are employed as a basis for standards derivation at this institution and that the basic underpinnings of the standard setting process may be related across models.

Further study of the derivation and application of standards in distance education evaluation will indicate whether the approaches employed at this institution pertains to distance education institutions in general. Further study may provide insight into the relationship between the application and derivation of standards across models and in terms of how theory and practical models apply to distance education evaluation in practice. It is recommended that this study be replicated at similar distance education institutions outside of Canada.

In the process of data collection and analysis at this institution, it was determined that many of the indicators or criteria were vague and considered difficult to measure. However there is little point in setting standards if measures to enable judgements regarding those standards are imprecise. It is recommended that a study, or series of studies, be undertaken to establish precise measures or indicators for a comprehensive set of course and program evaluation standards in distance education.

With the emergence of technology and greater demands for distance education in both industry and higher

education, further study of distance education program evaluation is warranted. Specifically, in this period of resource restraints and program cuts, the requirement to ensure education is at its best leads one to question how we are to conduct comprehensive program evaluation in the future. As decisions often rest on the results of evaluations, how standards are derived and applied in this process is critical to ensuring the validity of the findings. It is suggested that the scrutiny of the derivation and application and standards is warranted not only in distance education evaluation but in the evaluation of educational programs in general. It is only through further study in this area, that one can attempt to ascertain if the premises that our judgements rest upon are reliable and valid. It is recommended that similar studies be undertaken to explore evaluation standards and criteria in conventional institutions.

REFERENCES

Adams, M. (1991). Evaluation and Educational Performance: Making Sense of Performance Indicators in a Managerial Context. Journal of Further and Higher Education, 15 (1), p.3-17.

Alaska University. (1990). Evaluation in Distance Education at a Glance. Anchorage: Alaska University.

Allen, R. E. (1988). The Concise Oxford Dictionary of Current English, 7th ed, Oxford, Clarendon P.

Bartels, J. (1989). Methods and Experiences in Evaluation at the Fernuniversitat. Encuentro Iberoamericano de Educacion a Distancia, Costa Rica.

Bartels, J. (1991). Evaluation at the Fernuniversitat. In Schuemer, R.E. (ed.), Evaluation Concepts and Practices in Selected Distance Education Institutions. FernUniversitat, Hagen (West Germany). Zentrales Inst. fur Fernstudienforschung Arbeitsbereich

Birch, D. W. and Latcham, J. (1985). "Measuring College Performance". In Further Education Staff College: Assessing educational effectiveness and efficiency. Coombe Lodge Report No. 18(3). Blandon, the Further Education Staff College.

Birgitta, W. (1984). Strategies for Strengthening Student-Teacher Contact in Distance Education: Results of an evaluation of distance education in Swedish universities. DERG Papers No. 9. Open University (England), Distance Education Research Group.

Bloom, B. S. (1956). Taxonomy of Educational Objectives. New York: McKay.

Bloom, B. S. and Hastings, J. T. and Madaus, G. (1971). Handbook on Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill.

Capezio, P. and Morehouse, D. (1993). Taking the Mystery Out of TQM. NJ: Career Press.

Cizek, G. J. (1993). Reconsidering Standards and Criteria Journal of Educational Measurement, 30 (2), 93-106.

Coldeway, D. O. and Spencer, R. E. (1980). The Measurement of Attrition and Completion in Distance Learning Courses. REDEAL Technical Research Report, No. 8.

Coldeway, D. O. (1980). Distance Education from the Learners Perspective: Results of Individual Learner Tracking at Athabasca University. REDEAL Research Report, No. 10.

Coldeway, D. O. (1980). Factors Effecting Learner Motivation in Distance Education: The interaction between learner attributes and learner course performance. REDEAL Research Report, No.9.

Coldeway, D. O. (1986). Learner Characteristics and Success. In Mugridge, I. and Kaufman, D. (eds.) Distance Education in Canada. Dover, NH: Croom Helm.

Consortium for Policy Research in Education (1993). Developing Content Standards: Creating a Process for Change. NJ: Rutgers University.

Crawford, G. (1981). Student Completion Rates During Three Different Pacing Conditions. REDEAL Research Report, No. 12.

Crawford, G. and McGuire, S. (1994). Is it really worth it? Monitoring and formative evaluation of courses. A paper submitted to CADE' 94 ACDE.

Cross, P. K. (1993). Involving Faculty In TQM. Community College Journal, 63 (4), 15-18.

Danby, T. (1989).Evaluation and Standards. Unpublished paper, Memorial University of Newfoundland.

Daniel, J. S. and Snowden, B. L. (1981). The Management of Small Open Universities. In Pentz, M. J. and Neil, M. W. (eds.), Evaluation of Adults at a Distance. London: Kogan Page.

De Machado and Machado (1991). Institutional Evaluation of the National Open University of Venezuela: A Proposal. In Schuemer, R.E. (ed.), Evaluation Concepts and Practices in Selected Distance Education Institutions. FernUniversitat, Hagen (West Germany). Zentrales Inst. fur Fernstudienforschung Arbeitsbereich

Duffy, (1991). In Lewis, R.G. and Smith, D.H. (1994). Total Quality in Higher Education. Delray Beach, FL: St. Lucie Press.

Eisner, E. (1979). The Educational Imagination. New York: Macmillan.

Engineering Professor's Conference, (1992). Quality Assurance in Higher Education. Occasional Paper, No 4, EPC.

Evans, T. D. (1985). Just another brick in the wall: some reflections on student-based evaluation in distance education. Open Campus, 51-67.

Fadale, L. M. and Winter, G. M. (1985). Assessing the Effectiveness of Developmental/Remedial Programs in Two Year Colleges. Chicago: The Annual Meeting of the American Educational Research Association.

Fink, A. and Kosecoff, J. (1980). How to: Evaluate Education Programs. Arlington, VA: Capital Publications.

Ganor, M. (1991). Evaluation at a Distance. In Schuemer, R.E. (ed.), Evaluation Concepts and Practices in Selected Distance Education Institutions. FernUniversitat, Hagen (West Germany). Zentrales Inst. fur Fernstudienforschung Arbeitsbereich

Garrison, D. R. and Shale, D. (1987). Education at a Distance: From Issues to Practice. Malabar, Florida: R.E. Kreiger Publishing Co.

Gooler, D. D. (1979). Evaluating Distance Education Programs. Canadian Journal of Distance Education, 6(1) 43-55.

Guba, E. (1978). Toward a Methodology of Naturalistic Inquiry in Educational Evaluation. Los Angeles: Centre for the Study of Evaluation, UCLA.

Guba, E. and Lincoln, Y. (1981) Effective Evaluation. London: Jossey-Bass.

Guba, E. G. and Lincoln, Y. S. (1982). Effective Evaluation. San Francisco: Jossey-Bass.

Guttentag, M. (1973). Subjectivity and its use in evaluation research. Evaluation 1, 2: 60-65.

Harris, J. W. and Bagget, M. J. (1992). Quality Quest in the Academic Process. Birmingham, AL: Samford University.

Harrison, P. J., Seeman, B. J., Behm, R., Farhad, S., Molise, G. and Williams, D. (1991). Development of a distance education assessment instrument. Educational Technology Research and Development. 39 (4), p. 65-77.

Harvey, L. and Green, D. (1993). Defining Quality. Assessment and Evaluation in Higher Education. v18 n1 p9-34 1993

Holmberg, B. (1981). Trends and Status of Distance Education. London: Kogan Page.

Holmberg, B. (1986). Growth and Structure of Distance Education.Kent: Croom Helm.

Holmberg, B. (1989). Theory and Practice of Distance Education. Boston: Routeledge and Kegan Paul.

Holt, D. M. (1985). Theoretical Models of Evaluation. Open Campus, 3-20.

Hotchkis, R. D. and Driedger, L. (1990). Distance Education in Northern and Remote Communities: Understanding Social Networks, Change and Process. In Wall, D. and Owen, M. (eds), Distance Education and Sustainable Community Development. Edmonton, AB: Athabasca University Press.

House, E. R. (1980). Evaluating With Validity. CA: SAGE Publications.

House, E.R. (1980). Ch. 2. Taxonomy: Evaluating with Validity. Figure 1, p.23.

Hubbard, D. L. (1993). Continuous Quality Improvement: Making The Transition to Education. Marysville, MO: Prescott Publishing.

Ilyn, (1983). Cited in Moore, M. (1986). Self-Directed Learning and Distance Education. Journal of Distance Education, 1(1), pp 7-24.

Kaufman, R. (1991). Toward Total Quality "Plus". Training, v28 12, p50-54.

Kaye, A. (1988). Distance Education: The State of the Art. Prospects, 18 (1) 43-54.

Kaye, A. and Rumble, G. (1981). Distance Teaching for Higher and Adult Education. London: Croom Helm.

Keegan, D. (1986). The Foundations of Distance Education. London: Croom Helm.

Keegan, D. and Rumble, G. (1982). The DTUs: an appraisal. In Rumble, G. and Harry, K. (eds.) The distance teaching universities. London: Croom Helm, 222-250.

Kennedy, M. (1994). Setting Standards For Evaluating Distance Education Programs: Three Case Studies. Unpublished paper, Memorial University of Newfoundland.

Kennedy, M. (1993). A Study of Evaluation Standards within Distance Education Courses and Programs at Higher Education Institutions. Unpublished paper, Memorial University of Newfoundland.

Kirby, D. (1993). Is it time to rid universities of distance education? Journal of Distance Education, 8 (2), 69-72.

Knapper, K. C. and Cropley, J. (1985). Lifelong Learning and Higher Education. Kent: Croom Helm.

Koul, B. N. (1991). Evaluation at Indira Gandhi National Open University. In Schuemer, R.E. (ed.), Evaluation Concepts and Practices in Selected Distance Education Institutions. FernUniversitat, Hagen (West Germany). Zentrales Inst. fur Fernstudienforschung Arbeitsbereich.

Krippendorff, D. K. (1980). Analyzing Qualitative Data. San Francisco, CA: Jossey-Bass.

Lewis, R.G. and Smith, D.H. (1994). Total Quality in Higher Education. Delray Beach, FL: St. Lucie Press.

Lewey, A. (1983). Evaluation Standards: Comments from Israel. Montreal, QB: Annual Meeting of the American Educational Research Association, presentation paper.

Lincoln, Y. S. and Guba, E. G. (1985). Naturalistic Inquiry. Newbury Park, CA: Sage.

Lipsey, M. W. and Pollard, J. A. (1989). Driving toward theory in program evaluation: More models to choose from. Evaluation and Program Planning, 12 (4), 317-328.

MacDonald, B. (1974). Evaluation and the Control of Education. Norwich, England: Centre for Applied Research in Education.

Mager, R. F. (1962). Preparing Objectives for Programmed Instruction. San Fransico, CA: Fearon.

Markwood, R.A. and Johnstone, S.M. (1992). New Pathways to a Degree. Project Evaluation: First Year Report. Annenberg/CPB Project, Washington, DC.; Western Interstate Commission for Higher Education, Boulder, CO. Western Cooperative for Educational Communications.

McAnany, E. G. (1982). Distance Education: Evaluating New Approaches in Education for Developing Countries. Evaluation in Education: International Progress, 6 (3), 289.

McCaslin, N. L. and Headley, W. S. (1993). A national Study of Approved State Systems of Performance Measures and Standards for Vocational Education. Ohio State University, Columbus.

McIsaac, M. S. (1989). Problems Affecting Evaluation of Distance Education in Developing Countries. The Annual Meeting of the American Educational Research Association, 1989.

Merriam S. B. (1988). Case Study Research in Education. San Francisco, CA: Jossey-Bass.

Miller, C. D. and Clouse, R. W. (1994). Technology-Based Distance Learning: Present and Future Directions in Business and Education. Educational Technology Systems, 22 (3), 191.

Moore, M. (1986). Self-Directed Learning and Distance Education, Journal of Distance Education. 1(1). pp 7-24.

Moore, M. G. (1991). International Aspects of Independent Study. In Watkins, B.L. and Wright, S.J. (eds.), The Foundations of American Distance Education. Iowa: Kendall/Hunt Publishing.

Nevo, D. L. and Friedman, E. (1992). Relevance and Methodological Adequacy: A study of evaluation reports. San Francisco, CA: The Annual Meeting of the Educational Research Association.

New Zealand Qualifications Authority (1993). Quality Management Systems for the National Qualifications Framework: Quality Assurance in Education and Training. Wellington: New Zealand Qualifications Authority.

Orville, N. (1993). Conceptual Framework for Revising the Secondary Vocational Education Program Evaluation System. Stout Menomonie: Wisconsin University, Final Report.

Osbourne, B. (1993). What Education Leaders Need to Know Generally about Quality and Standards. Oklahoma: The Annual Creating the Quality School Conference, 2nd.

Parlett, M. and Hamilton, D. (1977). Evaluation as illumination: a new approach to the study of innovatory programmes. In D. Hamilton et al.(eds.) Beyond the Numbers Game. London: Macmillan.

Passow, H. A. (1987). Reporting the Results of Evaluation Studies. International Journal of Educational Research, 11 (1), 115-123.

Patton, M. Q. (1982). Practical Evaluation. Beverly Hills, CA: Sage.

Paul, R. (1990). Open Learning and Open Management. New York, NY: Nichols/GP Publishing.

Paul, R. (1993). Cited in Willis, B. (ed.) Distance Education: A Practical Guide. Englewood Cliffs, NJ: Educational Technology Publications.

Perry, W. (1977). The Open University. San Francisco, CA: Jossey Bass.

Perry, W. (1990). The genesis of the first open university. In Koul, B. N., and Jenkins J. (Eds.) Distance Education. London: Kogan Page.

Peruniak, G. (1980). Seminars as an Instructional Strategy in Distance Education. REDEAL Research Report, No. 8.

Popham, W. J. (1975). Educational Evaluation. Englewood Cliffs, NJ: Prentice-Hall.

Powell, R. and Coldeway, D. O. (1980). Effects of Self-Control Package on the Behaviour of the Distance Learner. REDEAL Research Report No. 8.

Pring, R. (1992). Standards and Quality in Education. British Journal of Education, 40(1), 4-22.

Provus, M. M. (1969). Evaluation of Ongoing Programs in the Public School System. In Tyler, R. W. (ed.), Educational Evaluation: New Roles, new means. Chicago: The 68th yearbook of the National Society for the Study of Education, Part Two.

Rumble, G. and Harry, K. (1982). The distance teaching universities. London: Croom Helm.

Rumble, G. (1986). The Planning and Management of Distance Education. Kent: Croom Helm.

Schall, K. and Stuart, K. (1993). Evaluating Interactive Video Teletraining (IVT). Journal of Instruction Delivery Systems, 7 (3), 6-10.

Schmoker, M. and Wilson, R. B. (1993). Adapting Total Quality Doesn't Mean Turning Learning into a Business. Educational Leadership, 51 (1), 62-63.

Scriven, M. (1967). The methodology of evaluation. In R.E. Stake (Ed.), Curriculum evaluation. American Educational Research Association Monograph series on evaluation, No. 1, Chicago: Rand McNally.

Scriven, M. (1972). Objectivity and Subjectivity in Educational Research. In L.G. Thomas (Ed.), Philosophical Redirection of Educational Research. Chicago: National Society for the Study of Education, University of Chicago Press.

Scriven, M. (1973). Goal free evaluation. In E.R. House (ed.) School Evaluation. Berkeley: McCutchan.

Scriven, M. (1991). Evaluation Thesaurus, (Fourth Ed.). Newbury Park, CA: Sage.

Sevick, C. (1993). Conceptual Framework for Revising the Secondary Vocational Education Program Evaluation System. Final Report. Wisconsin University, Stout Menomonie.

Shale, D. G. (1982). Attrition: A Case Study. In Daniel, Stroud and Thompson (eds.), Learning at a Distance: A World Perspective. Edmonton: Athabasca University, pp113-117.

Shale, D. G. (1984). A Study of Course Completion Rates. Institutional Studies Report, Vol. 6. Edmonton: Athabasca University.

Spencer, R. E. (1980). The Effects of Computer Generated Schedules on the Performance of Athabasca University Learners. REDEAL Research Report, # 5.

Stake, R. E. (1967). The countenance of educational evaluation. Teachers College Record, 68, 523-540

Stake, R. E. (1975). Evaluating the Arts in Education. Columbus, Ohio: Charles E. Merril.

Stake, R. E. (1991). Retrospective on "the Countenance of Educational Evaluation". In Stake, R.E. (ed.), Evaluation and Education: At Quarter Century.

Stake, R. E. and Denny, T. (1969). Needed Concepts and Techniques for Utilizing More Fully the Potential of Evaluation. In R.W. Tyler (Ed.), Educational Evaluation: New Roles, New Means. The 68th Yearbook of the National Society for the Study of Education, Part II. Chicago: National Society for the Study of Education, 1969.

Stufflebeam, D. L. (1973). An introduction to the PDK book. In B. Worthen and J. R. Saunders (eds.) Educational Evaluation: Theory and Practice. Worthington, OH: Charles A. Jones.

Sutcliffe, W. and Pollock, J. (1992). Can the Total Quality Management Approach Used in Industry be Transferred to Institutions of Higher Education? The Vocational Aspect of Education, 44 (1), 11-28.

Tannock, J. D. T. (1992). A New Approach to Quality Assurance for Higher Education. Higher Education Quarterly, 46 (1) 108-123

Thorpe, M. (1988). Evaluating Open and Distance Learning. Essex, G.B.: Longman.

Tyler, R. W. (1950) Basic Principles of Curriculum and Instruction. Chicago: Univ of Chicago Press.

Van Vught, F.A. and Westerheijden, D.F. (1994). Towards a General Model of Quality Assessment in Higher Education. Higher-Education, v28(3) pp 355-71.

Verduin, J. R. and Clark, T. A. (1991). Distance Education. San Francisco: Jossey-Bass.

Wagner, R. M. K. (1981). University Accommodation of Distance Education in Canada. Journal of Distance Education, 3 (1).

Wagner, E.D. (1993). Variables Affecting Distance Education Program Success. Educational Technology, April 1993.

Wedemyer, C. A. (1981). Learning at the Back Door. Madison, WI: The University of Wisconsin Press.

Weiss, C. H. (1991). Evaluation Research in the Political Context: Sixteen Years and Four Administrations Later. In Stake, R.E. (ed.), Evaluation and Education: At Quarter Century.

Willis, B. (1993). Distance Education: A Practical Guide. Englewood Cliffs, NJ: Educational Technology Publications.

Wolf, R. L. (1975). Trial by jury: the process. Phi Delta Kappan 57 (November): 185-187.

Woodley, A. (1991). Evaluation at the British Open University. In Schuemer, R.E. (ed.) Evaluation Concepts and Practices in Selected Distance Education Institutions. FernUniversitat, Hagen (West Germany). Zentrales Inst. fur Fernstudienforschung Arbeitsbereich.

Worthen, B. R. and Saunders, G. L (1973). Educational Evaluation: Theory and Practice. Worthington, Ohio: Charles, A. Jones Publishing Company.

APPENDIX A

DATA COLLECTION INSTRUMENTS

Evaluation Areas and Criteria/Indicators Interview Guide

Areas & Criteria	**Utilization**	**Evaluation Type**	**Measurement**	**Derivation**
Attendance Enrollment Criteria: completion taking future courses equal access				
Cost Criteria: per student per course				

Areas & Criteria	Utilization	Evaluation Type	Measurement	Derivation
Learning Criteria: geared to learners' level quantity learned: tests & exams quality learned: test, exam & instructor feedback graduate future job or education individual learning skills developed self-instructional: self evaluation opportunity				

Areas & Criteria	Utilization	Evaluation Type	Measurement	Derivation
Support Criteria: Technical access & availability Support from: organizational structure functions tutorial support, feedback, & timing				

Areas & Criteria	Utilization	Evaluation Type	Measurement	Derivation
Student Satisfaction Criteria: Course material: interesting essential updated adequate time comprehension presentation Appropriateness of: assigned work learning objectives grading audio-visual support				

Areas & Criteria	Utilization	Evaluation Type	Measurement	Derivation
Course Planning Criteria: need for the course utility of the course Economic viability: cost of materials & development validation of course materials prior to use				

QUESTIONNAIRE

INSTRUCTIONS

The following instructions apply to the questionnaire attached. Therefore, this page can be removed for easy reference. Listed in the questionnaire is a compilation of evaluation criteria, drawn from research literature, which are employed in distance education course and program evaluation. (All of them may or may not apply to evaluation conducted at this institution.) **For each criterion please indicate a response to the following questions in the column provided:**

Column 1: Do you employ this criterion when conducting any form of course or program evaluation? Indicate A or B.

a) Yes b) No (If no, please begin the next criteria.)

Column 2: If yes, what type of evaluation do you use this criterion for? Indicate A, B, C or D.

a) formal summative evaluation
b) formal formative evaluation
c) informal evaluation/monitoring
d) other (please specify in the comments section)

Column 3: What is an indication that a course or program is doing well in terms of this criterion? (i.e. How do you measure that this criterion has been successfully met?)

Column 4: Where does this measurement come from? (How was it derived?) Indicate A, B, C, D or E. Is it based on?

a) stakeholder group input b) university mandate
c) experience with courses d) research literature
e) other (please specify in the comments section)

To ensure that the intent of the questions are fully understood an example of completed responses, about a selected criterion, is provided below.

Criterion: COURSE COMPLETION

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?
A	C	50% of students, who initially register, complete the course.	D

Comments: _____

_____

Please indicate which of the following types of evaluation you conduct primarily in your field of specialization by circling a response.

a) program evaluation b) course evaluation
c) other (please specify below)

_____

_____

Below and on the following pages each evaluation criterion has been listed with a four column response table (as in the example on the previous page). Additional space has also been provided below each response table if you wish to elaborate or comment further on the criterion. Remember to refer to the page of Instructions, if required, when completing a response table.

Criterion: STUDENT COURSE COMPLETION

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments: _____

_____

Criterion: STUDENTS TAKING FUTURE COURSES

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

COST CRITERIA

Criterion: COST PER STUDENT

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: COST PER COURSE

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments: _____

_____

LEARNING CRITERIA

Criterion: COURSE AIMED AT THE LEARNER'S LEVEL

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments: _____

_____

Criterion: COURSE TEST AND EXAM SCORES

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: INDEPENDENT LEARNING SKILLS DEVELOPED

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: STUDENT EMPLOYMENT AFTER GRADUATING

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: NUMBER OF STUDENTS ACCEPTED TO GRADUATE SCHOOL

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: QUALITY OF LEARNING

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

SUPPORT CRITERIA

Criterion: ORGANIZATIONAL/ ADMINISTRATIVE SUPPORT

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: TUTORIAL SUPPORT

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: TUTORIAL FEEDBACK

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: TIMING OF FEEDBACK

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: LIBRARY SUPPORT

USED ?	TYPE ?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

STUDENT SATISFACTION CRITERIA

Criterion: STUDENT INTEREST IN COURSE MATERIAL

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments: _____

_____

Criterion: UP TO DATE COURSE MATERIAL

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments: _____

_____

Criterion: PRESENTATION OF COURSE MATERIALS

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: COURSE MATERIALS - ESSENTIAL IN NATURE

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: APPROPRIATENESS OF ASSIGNMENTS

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: APPROPRIATENESS OF LEARNING OBJECTIVES

USED?	TYPE ?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: APPROPRIATENESS OF GRADING

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: APPROPRIATENESS OF ANY COURSE AUDIO-VISUAL SUPPORT

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: ADEQUATE TIME TO COMPLETE COURSE MATERIAL

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments: _____

_____

Criterion: COMPREHENSION OF MATERIALS

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments: _____

_____

COURSE PLANNING CRITERIA

Criterion: ECONOMIC VIABILITY OF THE COURSE

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments: _____

_____

Criterion: FIELD TESTING OF COURSE MATERIALS

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments: _____

_____

Criterion: NEED FOR THE COURSE

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: UTILITY OF THE COURSE

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: COST OF MATERIALS AND DEVELOPMENT

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: LEVEL OF SELF-INSTRUCTION IN COURSE MATERIALS

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

Criterion: QUALITY OF TESTS AND EXAMS

USED?	TYPE?	WHAT IS THE MEASURE OF SUCCESS FOR THIS CRITERION?	HOW DERIVED? BASED ON WHAT?

Comments:_____

_____

THANK YOU FOR PARTICIPATING IN THIS STUDY!

APPENDIX B

CORRESPONDENCE

2736 Cyprus Ave. SW
Calgary, AB
T3E 7A2

Chairperson of the Ethics Review Committee
[redacted]

Dear [redacted]

I am a graduate student in the Faculty of Education at Memorial University and I am currently completing the thesis portion of the degree under the supervision of Dr. Mary Kennedy. I plan to conduct research at [redacted] to explore the evaluation of distance education programs. The focus of this study is the derivation and application of standards in the evaluation process and I will attempt to relate the findings to the evaluation literature. I am requesting your permission to examine documents and interview professional staff at [redacted]

Specifically, this study will involve: the examination of evaluation documentation and instruments; and, the conduct of semi-structured to unstructured interviews with program administrators/coordinators, instructional designers, program development team members, evaluation division personnel and course instructors. This portion of the study is expected to be complete in ten to fifteen working days, depending on the availability of personnel and any other limitations or unforeseen circumstances.

Participation in the study is completely voluntary. Any individual wishing to withdraw from the study may do so, without prejudice, at any time or may refrain from answering any questions that he or she prefers to omit. All information gathered in this study is completely confidential and at no time will individuals be identified. A cassette recorder will be employed during interviews, with participants' knowledge, in order to facilitate accuracy when analyzing responses to questions. Upon request, any individual's tape recording will be erased upon completion of the study. Recordings are privy only to the immediate researchers, Nadine Flannigan-Wheeler and Dr. Mary Kennedy. The research results will be available to participants, upon request, in the form of a thesis report once the study is concluded.

The researcher is not receiving any form of remuneration for this study. Furthermore, this study has received approval from the Faculty of Education's Ethics Review Committee. If you are in agreement with having the professional staff of [redacted] participate in this study please sign below and retain one copy for reference. If you have questions at any time and wish to speak with a resource person not associated with the study, please contact Dr. Patricia Canning, Associate Dean, Research and Development.

I would appreciate it if you could return this sheet to me as soon as possible. Thank you for your consideration of this request.

Yours sincerely,

N. Flannigan-Wheeler

I _____ the Chairperson of the Ethics Review Committee at [redacted] hereby give approval for the participation of professional staff and examination of evaluation documentation in the study of distance education evaluation undertaken by Nadine Flannigan-Wheeler. I understand that participation is entirely voluntary and that an I can withdraw permission at any time. All information is strictly confidential and no individual will be identified.

_____ _____

Date Signature

June 1, 1995

Ms. Nadine Flannigan-Wheeler
2736 Cyprus Ave. SW
Calgary, AB T3E 7A2

Dear Ms. Flannigan-Wheeler:

The [redacted] Human Subjects Committee has approved your plan to conduct the study entitled "The Application and Derivation of Standards in Distance Education Program Evaluation." The committee agreed that you plan meets our ethical criteria for research utilizing human subjects. In conducting the study, please ensure that you keep the completed consent forms on file for possible future reference by the committee.

Sincerely,

[redacted]

[redacted]
Associate Professor

cc. [redacted] Acting Vice-President Academic

2736 Cyprus Ave. SW
Calgary, AB
T3E 7A2

To whom it may concern,

I am a graduate student in the Faculty of Education at Memorial University and I am currently completing the thesis portion of the degree under the supervision of Dr. Mary Kennedy. I plan to conduct research at [redacted] to explore the evaluation of distance education programs. The focus of this study is the derivation and application of standards in the evaluation process and I will attempt to relate the findings to the evaluation literature. I am requesting your permission to interview you.

Specifically, this study will involve: the examination of evaluation documentation and instruments; and, the conduct of semi-structured to unstructured interviews with program administrators/coordinators, instructional designers, program development team members, evaluation division personnel and course instructors. This portion of the study is expected to be complete in ten to fifteen working days. One forty-five minute interview will be conducted and a second follow-up interview of fifteen minutes duration may be required.

Participation in the study is completely voluntary. Any individual wishing to withdraw from the study may do so, without prejudice, at any time or may refrain from answering any questions that he or she prefers to omit. All information gathered in this study is completely confidential and at no time will individuals or the institution be identified. A cassette recorder will be employed during interviews, with participants' knowledge, in order to facilitate accuracy when analyzing responses to questions. Upon request, any individual's tape recording will be erased upon completion of the study. Recordings are privy only to the immediate researchers, Nadine Flannigan-Wheeler and Dr. Mary Kennedy. The research results will be available to participants, upon request, in the form of a thesis report once the study is concluded.

The researcher is not receiving any form of remuneration for this study. Furthermore, this study has received approval from the Faculty of Education's Ethics Review Committee. If you are willing to participate in this study please sign below and retain one copy for reference. If you have questions at any time and wish to speak with a resource person not associated with the study, please contact Dr. Stephen Norris, Associate Dean, Research and Development.

I would appreciate it if you could return this sheet to me as soon as possible. Thank you for your consideration of this request.

Yours sincerely,

Nadine Flannigan-Wheeler

I_____hereby give my approval to participate in the study of distance education evaluation undertaken by Nadine Flannigan-Wheeler. I understand that participation is entirely voluntary and that I can withdraw permission at any time. All information is strictly confidential and no individual nor [redacted] will be identified.

_____ _____

Date Signature

34 Summerwind Cres.
Napean, ON
K2G 6G5

I am a graduate student in the Faculty of Education at Memorial University and I am currently completing the thesis portion of the degree under the supervision of Dr. Mary Kennedy. I have been conducting research at [redacted] to explore the evaluation of distance education programs. The focus of this study is the derivation and application of standards in the evaluation process and I will attempt to relate the findings to the evaluation literature.

Specifically, this study involves: the examination of evaluation documentation and instruments; and, the conduct of semi-structured to unstructured interviews with program administrators/coordinators, instructional designers, program development team members, evaluation division personnel and course instructors.

Due to time constraints, I did not get the opportunity to interview you in person. However, you were recommended to me, as someone who can contribute to this study, by your colleagues that I did interview. Therefore, I am sending this questionnaire and hoping that you will take the time to complete it. It may appear quite detailed at first glance, but is actually based on four recurring questions which will not take long to answer once you get going. In addition, writing has been minimized and most responses require that you simply write a letter.

Participation in the study is completely voluntary. Any individual wishing to withdraw from the study may do so, without prejudice, at any time or may refrain from answering any questions that he or she prefers to omit. All information gathered in this study is completely confidential and at no time will individuals or the institution be identified.. Completed questionnaires are privy only to the immediate researcher Nadine Flannigan-Wheeler. The research results will be available to participants, upon request, in the form of a thesis report once the study is concluded.

The researcher is not receiving any form of remuneration for this study. Furthermore, this study has received approval from the Faculty of Education's Ethics Review Committee at Memorial University and at [redacted] If you have questions at any time and wish to speak with a resource person not associated with the study, please contact Dr. Stephen Norris, Associate Dean, Research and Development.

I will follow up with a phone call, in late August, to see if you have any questions about the questionnaire. If you are on vacation or working elsewhere at that time I will try to contact you at a later date. This does not pose a problem as the deadline for returning questionnaires is November 1st.

Thank you for your consideration of this request. Enclosed is the questionnaire and a stamped, self addressed envelope. I look forward to hearing from you.

Yours sincerely,

Nadine Flannigan-Wheeler

Nadine Flannigan-Wheeler