Factors Affecting Program Evaluation Behaviours of Natural Resource Extension Practitioners in the United States

by

SHAWN MORFORD

A THESIS SUBMITTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

Faculty of Forestry, Department of Forest Resources Management

We accept this thesis as conforming
to the required standard

The University of British Columbia

September 2004

Factors Affecting Program Evaluation Behaviours of Natural Resource Extension Practitioners in the United States

"The future of Extension depends on its ability to document its impact and demonstrate its effectiveness" (Warner and Christenson 1984)

ABSTRACT

The systematic process of measuring and reporting on impacts of publicly funded programs, referred to as "program evaluation," is receiving increased attention among federal, provincial/state, and local organizations throughout North America, including education programs funded by the U.S. Department of Agriculture's Cooperative State Research and Extension Education Service (CSREES). Extension professionals who deliver these programs are expected to collect evaluation data that show how programs change knowledge, skills, attitudes, and behaviours of clientele as well as longer-term impacts of interventions. Despite the demand for accountability of Extension programs, the frequency and level of program evaluation conducted by natural resource Extension professionals are highly variable across individuals and states.

Using a survey, this study investigated motivation, attitude, perceived organizational commitment, and personal characteristics of natural resource extension professionals (NREPs) as they related to the level of program evaluation conducted. Using Analysis of Variance, Multiple Linear Regression and Cross tabulations, this study examined program evaluation behaviours and attitudes of NREPs who work under the CSREES system in the United States in fields such as watershed stewardship, forest products marketing, land use planning, flood mitigation, and forestry. The study examined how factors such as position classification, tenure status, personnel appraisal criteria, perceived organizational commitment, years of experience, and other factors influence program evaluation behaviour.

The study showed that age, years of experience in Extension, belief that performance appraisal is linked to evaluation behaviour, and attitude about program evaluation influence the level of evaluation conducted by NREPs. Position classification (i.e tenure-track/non-tenure track), access to evaluation specialists, perceived organizational commitment to evaluation, access to evaluation specialists, and confidence levels in conducting evaluation also may be linked with level of evaluation conducted but results are less conclusive for those variables. Recommendations are offered to improve the amount and quality of program evaluation. The results will contribute to the body of knowledge that will enhance the evaluation capacity within organizations, especially Extension, and ultimately lead to more effective, efficient, and well-funded natural resource Extension programming.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ACKNOWLEDGEMENTS

I owe a great debt of thanks to my supervisor, Dr. John Innes, for his guidance and encouragement during all stages of my doctoral work. I also thank other members of my committee, particularly Dr. Rob Kozak who coached me through the statistical analyses and Dr. Murari Suvedi who provided topical guidance and input throughout the process. I also thank my other committee members Dr. John McLean and Dr. George Hoberg for their time, encouragement, and input, and the Faculty of Forestry and the Ralph and Elizabeth Cochran family for support funds. I dedicate this work to my parents, Les and Vesta Morford, who have provided love and stability, and instilled me with a love for education and knowledge.

1. INTRODUCTION

1.1 Purpose

For the past 90 years, citizens and residents in the United States have participated in education programs through state land-grant universities that have a mandate to provide educational outreach to citizens. Organized by a national organization now called the Cooperative State Research and Extension Education Service (CSREES), the system employs outreach educators who provide non-formal education programs in fields such as agriculture, natural resources, home economics, and youth development in every state and U.S. territory. Like other government-supported organizations, managers of these programs are coming under increasing pressure to demonstrate impacts on indicators of quality of life, health, economic prosperity and environmental protection. Extension educators are expected not only to conduct high-quality programs, but also to evaluate them and report their impacts.

As I worked as an Extension professional in Oregon during the 1990s and more recently as a natural resource Extension professional (NREP) in Canada, I have been expected to evaluate my programs. Through the years, I have observed many colleagues who have regularly conducted program evaluation, but also knew those who did little or no evaluation of their programs. I did not have formal training in program evaluation; my attempts at evaluation of programs beyond the "end-of-event" questionnaires have historically been weak. In launching this study, I was interested in identifying what makes Extension practitioners in the U.S. system "tick" when it comes to evaluation. Why do some Extension professionals do evaluations and others do few? What can administrators do to induce more frequent and higher-level evaluations, if anything? Is it a matter of personal attitude, perceived commitment of the organization to evaluation, or promotion and tenure? Does it matter whether an Extension professional has a tenure-track position? Does it matter whether an Extension professional's salary is funded by grant dollars?

Given these questions, the purpose of this study was to investigate the factors influencing the level of program evaluation conducted by field and campus-based natural resource Extension practitioners in the U.S.A. The study assessed:

- attitude toward program evaluation;
- perceived organizational commitment to evaluation;
- level of evaluation normally conducted by individual NREPs, as measured on Bennett's Hierarchy; and
- NREP characteristics including age, years in the profession, funding source for position (grant/core-funded), position classification (tenure track/non-tenure track), job classification (campus-based specialist, county/region-based agent, professional staff), tenure status, and the relationship of these factors to the level of evaluation conducted.

A secondary purpose of the research was to identify and describe knowledge gaps, and perceived barriers regarding program evaluation among NREPs to help administrators reduce barriers.

In Oregon, my position was 100% grant-funded and I was a non-tenure-track faculty member with the Oregon State University (OSU) Extension Service. The federal agency (U.S. Department of Agriculture) that funded my program required regular program evaluation results, and my OSU personnel appraisal also included criteria related to program evaluation.

This study also asks how land-grant universities that employ NREPs can increase their evaluation capacity overall. How can this study advance the emerging field of Evaluation Capacity Building (ECB) by identifying individual motivation and behaviours as structures within a framework of ECB?

1.2 Definitions

In this study, the following definitions apply:

a. Core competency: Knowledge, skills or abilities required of the job (Cooper and Graham 2001).

b. Evaluation Capacity Building (ECB): The intentional work to continuously create and sustain overall organizational processes that make quality evaluation and its use routine (Compton et al. 2002)

c. Extension: A non-formal educational process that involves a two-way information flow between knowledge centres (in this case, universities) and knowledge users (in this case, private forest land owners and other publics). This definition differs from its counterpart, "technology transfer" in that users of information are not passive receivers of information, but are active in defining and communicating their information needs to knowledge centres. The term, as used in this thesis, also refers to a profession, a university major (usually a graduate degree), and an organization (the word is short for "Cooperative State Research and Extension Education Service" in the U.S.).

d. Land grant universities: Universities in the U.S. established through the Morrill Act of 1862 by a land grant from the federal government to each state. Land was given with a directive to establish a college to extend research-based, practical methodologies outside the campus to develop an "informed, vigorous citizenry" (Sanderson 1988). Today, each state has a land grant university that offers research, teaching, and Extension services. The universities (for example, Washington State University, and the University of Wyoming) serve as the home for the Extension Service in each state. Some land grant universities (e.g. Oregon State University and Pennsylvania State University) have large natural resource Extension programs and faculty.

e. Natural resource Extension practitioner (NREP): Person employed as a full-time or part time Extension faculty or staff member by land grant universities whose mandate is to provide non-formal, research-based education regarding natural resources to various audiences and to communicate information gaps to researchers. In Colorado, the Extension function has been transferred by agreement to state forestry departments, so this definition includes those staff members.

f. Tenure-track NREP position: Type of Extension position established by some land grant universities in which the Extension professional is considered a regular faculty member of the university subject to the same academic expectations and privileges as other faculty members.

In some states, however, Extension practitioner positions are considered as "non-tenure-track" positions.

g. Program evaluation: the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about a program, improve effectiveness, and/or inform decisions about future programming (Patton 1997).

2.0 EXTENSION ACTIVITIES AND THEIR EVALUATION

In this chapter, I provide background on Extension as profession in the U.S. and discuss how demand for evaluation has increased in the profession. I will contrast this with a description of Extension and evaluation practices in Europe and British Columbia. I will describe previous studies that examined attitudes and behaviours of American Extension professionals towards evaluation and show where there are gaps in knowledge about their attitudes and behaviours. I will introduce the literature on "evaluation capacity building" (building organizational processes to ensure high-quality evaluations within an organization) as well as describe a planning and evaluation tool used in Extension (Bennett's Hierarchy) that I will use in my analysis later in the thesis. Finally, I introduce models of human and work motivation that serve as a theoretical foundation for my hypotheses.

2.1 Background on Extension in the United States

Extension in the U.S. owes its beginning to the philosophies of early American presidents George Washington (1789-1797) and Thomas Jefferson (1801-1809), who believed that strengthening agriculture was a key to developing the nation state. They expressed a desire to bring education to "common people, whose lives would be spent in the nation's businesses and trades," and who were largely rural and agrarian at the time (Sanderson 1988). The 1862 Morrill Act led to the establishment of the land-grant university system that featured agricultural research and education. One so-called "land-grant university" was established in each state; each included a mandate to extend research-based information to rural people to promote progress, prosperity, and democracy, in addition to teaching and research. Congressional district agriculture schools in Virginia were established in 1908 to train secondary students in agriculture and home economics and are cited for their foundational role in the work of the Extension Service in that state (Sutphin and Hillison 1999). The 1914 Smith Lever Act led to the creation of what is now known as the Cooperative State Research and Extension Education Service (CSREES), a unique partnership between county, state, and federal governments to extend research-based information to citizens. The term "cooperative" referred to the shared funding relationship between county, state, and federal governments.

The Extension "movement" was given a boost when Professor Seaman Knapp from Iowa State University enlisted thousands of farmers to establish demonstration farms in the early 1900s to exhibit advanced farming methods; the events were credited with the control of the otherwise devastating cotton boll weevil in the southern U.S. in 1904. The establishment of the Cooperative Extension Service followed, with its system of agricultural agents located in rural communities across the country. During World War II, the pressure to increase agricultural productivity led to the organization growing to over 11,000 staff members, with agents in nearly every county in the country.

While serving farmers was originally at the heart of the original Cooperative Extension Service, it has since broadened to include other audiences such as homemakers, youth, community leaders, sawmill owners, private forestland owners, and the public. In 1997, the CSREES system employed approximately 14,890 full time equivalents or FTEs (i.e. the equivalent of 14,890 full time positions) nationwide (Ahearn et al. 2003). Extension agents located in county and campus-based offices provided non-formal education programs in home economics, agriculture, rural economic development, fisheries, 4-H/Youth, and natural resources.

Table 2-1 is a list of 76 Land-Grant institutions in the U.S. and U.S. Territories.

Table 2-1 Land-Grant institutions in the U.S. and U.S. Territories

(source: National Association of State Universities and Land Grant Colleges. Available at http://www.nasulgc.org/about_Nasulgc/members_land_grant.htm).

ALABAMA
Alabama A&M University
Auburn University
Tuskegee University

AMERICAN SAMOA
American Samoa Community College

ARKANSAS
University of Arkansas, Fayetteville
University of Arkansas at Pine Bluff

ALASKA
University of Alaska System

ARIZONA
University of Arizona

CALIFORNIA
University of California

COLORADO
Colorado State University

CONNECTICUT
Connecticut Agricultural Experiment Station
University of Connecticut

DELAWARE
Delaware State University
University of Delaware

DISTRICT OF COLUMBIA
University of the District of Columbia

FLORIDA
Florida A&M University
University of Florida

MISSOURI
Lincoln University
University of Missouri System

MONTANA
Montana State University

NEBRASKA
University of Nebraska

NEVADA
University of Nevada, Reno

NEW HAMPSHIRE
University of New Hampshire

NEW JERSEY
Rutgers, The State University of New Jersey

NEW MEXICO
New Mexico State University

NEW YORK
Cornell University

NORTH CAROLINA
North Carolina A&T State University
North Carolina State University

NORTH DAKOTA
North Dakota State University

OHIO
The Ohio State University

OKLAHOMA
Langston University
Oklahoma State University

GEORGIA
Fort Valley State University
University of Georgia

GUAM
University of Guam

HAWAII
University of Hawaii

IDAHO
University of Idaho

ILLINOIS
University of Illinois

INDIANA
Purdue University

IOWA
Iowa State University

KANSAS
Kansas State University

KENTUCKY
Kentucky State University
University of Kentucky

LOUISIANA
Louisiana State University System
Southern University and A&M College

MAINE
University of Maine

MARYLAND
University of Maryland Eastern Shore
University of Maryland, College Park

MASSACHUSETTS
Massachusetts Institute of Technology
University of Massachusetts, Amherst

MICHIGAN
Michigan State University

MINNESOTA
University of Minnesota

MISSISSIPPI
Alcorn State University
Mississippi State University

OREGON
Oregon State University

PENNSYLVANIA
The Pennsylvania State University

PUERTO RICO
University of Puerto Rico Mayaguez

RHODE ISLAND
University of Rhode Island

SOUTH CAROLINA
Clemson University
South Carolina State University

SOUTH DAKOTA
South Dakota State University

TENNESSEE
Tennessee State University
University of Tennessee

TEXAS
Prairie View A&M University
Texas A&M University

UTAH
Utah State University

VERMONT
University of Vermont

VIRGIN ISLANDS
University of the Virgin Islands

VIRGINIA
American Indian Higher Education Consortium
Virginia Polytechnic Institute & State University
Virginia State University

WASHINGTON
Washington State University

WEST VIRGINIA
West Virginia University
West Virginia State University

WISCONSIN
University of Wisconsin-Madison

WYOMING
University of Wyoming

Social and economic change during the 1980s led the Cooperative Extension Service to broaden its mission and programs to more issue-based public education on a wider range of topics, particularly issues related to water (Ahearn et al. 2003). Major societal shifts such as globalization, the changing urban/rural interface, changing family structures, and increased information and communications technology led the Extension system to re-examine its mission and focus more on helping communities and citizens manage change (Kellogg Commission 2001).

The "Extension in the 1980s" report recommended that Extension "provide educational programs for improved decision making in the management of our natural resources and environment" (Sanderson 1988). Four Rural Development Centers were established by a Rural Development Act to link research with rural resource-based communities to focus on rural development issues during this era. By the mid-1990s, nearly a quarter of Extension positions were devoted to national initiatives that included a focus on water quality, sustainable agriculture, and communities in transition. In 1998 there were approximately 268 FTE natural resources Extension positions in the Extension system nationwide (Hamilton and Biles 1998). That number increased over the next five years as the definition of natural resources expanded to include areas such as water, wildlife, and other non-timber resources.

Historically called "agents" (some states are no longer using that title), county/region-based NREPs are employed by land grant universities (except in Colorado where natural resource Extension practitioners are linked to the state forestry agency) and are traditionally housed in county government offices to provide close links with citizens (Hamilton and Biles 1998). The classic model shows campus-based Extension "specialists" linking local Extension agents with university research. Campus-based specialists often have some portion of their position in research and teaching in addition to their Extension functions. Through linkages between citizens, county-based Extension agents, campus-based Extension specialists, and university researchers, citizens are tied to information that is "relevant and timely to put practical knowledge to work," as shown in Figure 2-1 (Hamilton and Biles 1998). Extension practitioners are responsible for planning and implementing educational programs, assessing local knowledge gaps, setting educational priorities, and feeding research needs back to university research installations and researchers. In many states, county governments contribute office space and funds for program assistant and clerical staff salaries. In some cases, Extension specialists link directly to citizens, depending on the topic and the audience.

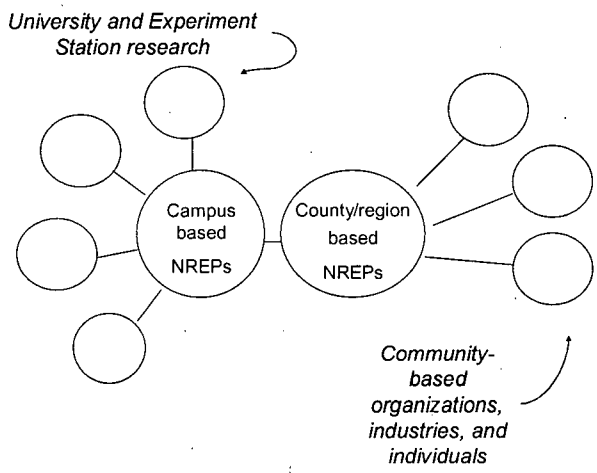

Figure 2-1 Classic Extension model showing linkages between university research, campus-based specialists, county/region-based agents, and citizens.

Much of the funding for natural resource Extension comes through the federal Renewable Resources Extension Act of 1978 (RREA 1978). Under the act, federal funds are provided for programs to address the needs of private forest and range land in the areas of rural and urban forest management, forest products marketing and utilization, range, wildlife, and fisheries management, outdoor recreation, and continuing education. Funds are allocated to states using a formula that considers the area of private forest and range land, wood industry employment, total state population, and urban population. Hamilton and Biles (1998) estimated that natural resource Extension budgets constituted approximately two percent or $28 million of the total annual CSREES budget. Agriculture and Natural Resources is the largest program area in the land-grant system. In some states such as Oregon, NREPs are considered tenure-track faculty, while in other states such as Washington they are considered non-tenure track faculty. Some funds for positions come from grants ("soft funds"), while other positions are core-funded, depending on the state.

2.2 The mandate for program evaluation

Several authors (Wandersman et al. 2003, Milstein et al. 2002, and Forest et al. 1989) refer to increasing federal and state legislative requirements for evaluation and accountability in the U.S.. The CSREES is among an increasing number of organizations expected to demonstrate the impacts of its programs (Taylor-Powell 2002, Forest et al. 1989). Despite its lengthy 90-year history, however, the CSREES has had a patchy history of program evaluation (Warner and Christenson 1984). The national CSREES office in Washington D.C. has recognized the need for improvement and has urged program evaluation at every level of the organization. During the winter and spring of 2002, members of the Extension Education Evaluation Topical Interest Group of the American Evaluation Association discussed how to induce an "evaluation

culture" within the CSREES at all levels of the organization, asserting that the responsibility for accountability rests with local-level practitioners as well as with higher-level administrators. Many state Extension offices have begun discussions and initiatives to increase the enculturation of evaluation in their organizations as a result of these discussions and other factors.

The U.S. Department of Agriculture (USDA) and land-grant institutions have been required to focus more closely on accountability, as required by the federal Government Performance and Results Act of 1993 (Ahearn et al. 2003 and AREERA 1998). As a result of the Act, the USDA created five goal areas and asked states to report on performance on these goals: a safe, secure food and fibre system, a healthy, well-nourished population, greater harmony between agriculture and the environment, enhanced economic opportunity, and quality of life. The Agriculture Research, Extension and Education Reform Act of 1998 (AREERA) required development of performance-based management and indicators related to outcomes such as changes in behaviours that result in an improvement in quality of life, quality of the environment, and level of stakeholder involvement in planning and evaluation. USDA programs such as the Sustainable Agriculture Research and Extension Project (SARE) have sparked new frameworks for evaluating complex sets of indicators (Suvedi et al. 2003).

In 2002, the CSREES created a Planning and Accountability Unit at the national headquarters in Washington D.C. to meet the agency's increasing demand for evaluation and planning resulting from recent federal legislation. The unit houses 11 staff members, including an economist, evaluation and program specialists, social science analysts, and program assistants. The unit's staff works with leaders of land-grant universities to meet planning and accountability requirements outlined by the Government Performance and Results Act, the AREERA and the President's Management Agenda (PMA), announced in 2001. Staff of this unit provides national guidelines for strategic planning and evaluation for state land-grant universities.

A 1982 study of 13,000 U.S. Extension faculty and staff, cited by Forest et al. (1989), showed that over 60% of Extension staff and administrators believed that Extension needed to be more accountable and provide more impact data. Between 63 and 84 percent (depending on the category) of respondents said that the organization should "do more" to measure program impact, report to major organizations, and inform politicians and the public. However, only 39% of staff said that there should be a formal state and national accountability system.

Forest et al. (1989) recommended that Extension practitioners become more pro-active in providing impact data, develop measurable outcome-based objectives for their programs, accept that systematic data collection on impacts is part of the accountability process, involve stakeholders in analysing evaluation data, and use evaluation data to improve existing programs. Despite these recommendations, evaluation capacity continues to be lacking among Extension personnel and training is believed to be one mechanism to bring

Extension personnel up to an acceptable level of competency in program evaluation (Taylor-Powell 2002). Current national guidelines of the CSREES call for improved continuing education for Extension staff on "methods and procedures essential for evaluating Extension programs." The guidelines state that Extension practitioners need educational experiences to assist them in "improving personal competencies in program analysis, interpretation, and accountability" (U.S. Department of Agriculture 2001) in order to improve the level of accountability of the Extension Service overall.

Bush et al. (1995) urged Extension professionals to not only collect impact data, but to examine the processes occurring between Extension activities and inputs and outcomes to understand why certain interventions are effective, not just that they "are." The Kellogg Commission cited the need to examine faculty promotion and tenure guidelines to ensure they recognize and reward faculty accountability (Kellogg Commission 2001). National guidelines presented by the Extension Committee on Organization and Policy recommended that in-service education for Extension professionals include training to increase skills in program development and analysis and accountability (ECOP(a) 2000 and ECOP(b) 2000).

Developing evaluation capacity within organizations has become a focus of many evaluation professionals outside as well as inside the CSREES. The theme of the 2001 American Evaluation Association annual conference, "Mainstreaming Evaluation," highlighted this trend. Literature such as Building Evaluation Capacity: 72 Activities for Teaching and Training by Preskill and Russ-Eft (2004), Building Evaluation Capacity: Lessons from Practice by Doyle and Lemaire (1999) and "The Mainstreaming of Evaluation" edited by Barnette and Sanders (2003) is surfacing. Some Extension evaluation specialists believe that increasing "evaluative thinking" (a conscious willingness to incorporate concepts of evaluation in programs and projects at multiple levels within an organization) is needed (Taylor-Powell 2002).

Extension administrators in the U.S. use a range of methods to encourage Extension practitioners to conduct evaluations of their programs, including training, hiring campus-based evaluation specialists to advise and assist practitioners, using program evaluation experience as a hiring criterion, and using program evaluation practices as a criterion in annual performance appraisals. However, many organizations have limited capacity for conducting program evaluations, lacking in-house skills, resources, and employee incentives for reliable, meaningful evaluations of their programs (Doyle and Lemaire 1999). If evaluation is to become a 'part of doing business' among practitioners within Extension (Decker 1990), then understanding what leads Extension faculty and staff to conduct – or not conduct - evaluations is a critical step.

2.3 Accountability in British Columbia

Government and other organizations in British Columbia have been subject to similar growing pressures for accountability as in the U.S., particularly since the mid-1990s (McDavid 2001).

The British Columbia Budget Transparency and Accountability Act of 2000 requires provincial agencies to establish evaluation and reporting systems to ensure accountability (BTAA 2000). The "New Era of Sustainable Forestry" introduced by the Liberal government in 2001 included provisions to increase accountability of forest licensees through a revised Forest and Range Practices Act in British Columbia (Hoberg 2002, Ministry of Forests 2004a). The new Act required the establishment of monitoring and evaluation systems to determine whether or not the desired results of the Act have been reached. In 2003, the Forest Practices Branch of the provincial Ministry of Forests launched an evaluation training pilot and released an evaluation plan for the revised Forest Practices and Range Act (Ministry of Forests 2004b).

The terms "results-based" and "accountability" have become part of the daily lexicon of many government and government-funded organizational leaders in British Columbia and elsewhere (Suvedi and Morford 2003). Interviews with provincial agency representatives in 2001 revealed that there is a strong belief that strategic planning and performance measurement have become a permanent feature of the administrative and political landscape in B.C. (McDavid 2001).

2.4 Extension and evaluation capacity in European and other countries

There is considerable literature related to forestry and natural resource Extension in Western Europe where there is a long history of extending information to private non-industrial forest land owners through a wide range of public and private organizations (Andersen 1997, G. Struhkamp, European Evaluation Association member, pers. comm., Nov. 2002). Forestry Extension services in the Nordic countries of Finland, Denmark, Sweden and Norway are provided through forestry societies (such as the Danish and Norwegian forestry societies), county forestry boards (in Sweden), and forest owner associations (particularly in Norway and Finland). Publicly funded and private forestry Extension services often work side-by-side and landowners have choices of services and approaches. In France, Extension services are provided by regional Centres of Forest Ownership (public organizations administered by elected representatives of private forest owners), a Forest Development Institute, local chambers of agriculture and private consultants (FAO 1997). In Bavaria, there are approximately 173 forest owner associations (FAO 1997).

In Great Britain, forestry ownership is characterized by small farm woodlands and a gradual change from predominantly state ownership to private ownership. The change from timber production to recreational uses has enabled Extension groups such as the Farming and Wildlife Advisory Group (FWAG), the Forestry Authority, and Game Conservancy Trust to have more significant roles with landowners and farmers. Extension foresters with FWAG provide farmers, landowners and other clients with the opportunities for environmental gain through cost effective, quality solutions to environmental issues. The FWAG extension staff provides farmers and landowners with technical advice on farm operation modifications and farm

feature enhancement to support wildlife, landscape, archaeology, access and other conservation values (FWAG 2004). The Game Conservancy Trust (GCT) provides advisory services for farmers and landowners on improving wildlife habitat. The GCT website describes services such as designing and enhancing woodlands, trout stream restoration, hedgerow and field boundary management, and designing seed crops for wintering birds (GCT 2004).

A Food and Agriculture Organization of the United Nations (FAO) survey in 1997 inventoried forestry Extension capacities in nine countries of Central and Eastern Europe (FAO 1997). Prompted by the increasing number of private forest owners due to tenure reform, education of new private forest land owners has become a priority for the governments of Armenia, the Czech Republic, Poland, Latvia, Croatia, Hungary, Romania, Slovakia and Slovenia. Latvia, for example, experienced an increase in private forest land ownership from 0% to 18% of the total forest land over a 10-year period between 1985 and 1995. However, as of 1997, staff of the State Forest Service had received no formal training in the provision of forestry Extension to private forest land owners. The survey indicated that the State Forest Service in Latvia planned to design and implement a comprehensive Extension system in the future.

Responsibility for forestry Extension in most Central and Eastern European countries is held by state agriculture or forestry organizations and the audience in most countries (except for Croatia, Poland and Slovenia where private forest ownership had previously existed) have not had exposure to forestry. Many are absentee owners who live in urban areas. Landowner associations are rare in Central and Eastern Europe (FAO 1997). The survey revealed that no formal Extension evaluation mechanisms or criteria had been established in Armenia but evaluation questions had been identified. Other country profiles in the study did not discuss Extension evaluation.

Evaluation has been described as a "young field" in Europe by comparison to North America (G. Struhkamp, European Evaluation Association member, pers. comm., Nov. 2002). For example the European Evaluation Association was formed in 1994 and the German Society for Evaluation was founded in 1997, compared with the American Evaluation Association in 1986 and the Canadian Evaluation Society in 1981. The American Evaluation Association website lists 22 national evaluation associations, including those as diverse as the Nigerian Network of Monitoring and Evaluation, the African Evaluation Association (Afrea 2004), the Israeli Association for Program Evaluation (IAPE 2004), and the Russian International Project Evaluation Society (IPEN 2004). The website of the International Energy Program Evaluation conference lists contacts for 27 additional evaluation associations from countries such as Zimbabwe, Thailand, Egypt, Nepal and Brazil (IEPEC 2004). The International Organisation for Cooperation in Evaluation was launched in 2003 as an umbrella organization involving evaluation bodies from Australasia, North America, Europe, and Latin America (IOCE 2004).

Table 2-2 Websites of international evaluation associations

International Evaluation Associations 2004	Web Sites
International Organization for Cooperation in Evaluation	http://www.internationalevaluation.com/
African Evaluation Association	http://www.afrea.org
Australasian Evaluation Society	http://www.parklane.com.au/aes/
Canada--Association pour le Développement des Méthodologies d'Evaluation en Education	http://www.uottawa.ca/associations/admee/
Canadian Evaluation Society	http://www.evaluationcanada.ca/
Europe--Association pour le Développement des Méthodologies d'Evaluation en Education- in French	http://www.irdp.ch/admee/
European Evaluation Society	http://www.europeanevaluation.org/
French Evaluation Association	http://www.sfe.asso.fr/
German Center for Evaluation in German	http://www.uni-koeln.de/ew-fak/Wiso/
Italian Evaluation Society in Italian	http://www.valutazione.it/
Russian International Project Evaluation Network (IPEN) in Russian	http://ipen21.org/ipen/
Swiss Evaluation Society in Swiss	http://www.seval.ch/de/index.cfm
UK Evaluation Society: Evaluators' Web	http://www.evaluation.org.uk/

The International Development Evaluation Association (IDEAS) was founded in 2002 in response to an acknowledged need to improve the practice of evaluation among international development organizations and developing countries (IDEAS 2004). The Overseas Development Institute, headquartered in the U.K., launched the Neuchâtel Initiative to bring together an informal group of bilateral and multilateral donor agencies in 1995 to develop a shared vision on how to address challenges and changes facing agricultural Extension in sub-Saharan Africa. The initiative included creation of evaluation guidelines, outlined in a publication called "*Guide for Monitoring, Evaluation and Joint Analyses of Pluralistic Extension Support*" (ODI 2004). While there is a significant literature on natural resource Extension as well as evaluation networks and organizations in industrialized and emerging countries in Europe, there is considerably less written about Extension evaluation *per se*. Correspondence with forestry Extension colleagues in Slovenia and Denmark reveals that there are few processes or policies that have been established to systematically evaluate forestry Extension services in those countries and most evaluations simply involve end-of-event questionnaires designed to improve teaching performance (J. Begus, Slovenia Forest Service, pers. comm., June 2004). Discussion with an Extension academic and practitioner in Bavaria, Germany reinforced the view that most forestry Extension services in Western Europe are demand-

driven (driven by the demand of the audience) and that there is a lack of pro-active planning and development of measurable objectives or even identification of the target audience. R. Beck (Lehrstuhl für Forstpolitik und Forstgeschichte, Technical University of Munich, Germany, pers. comm., June 2004) describes evaluation in forestry Extension in Western Europe as a matter of record keeping of money spent, labour invested, and number and type of courses, number of attendees and number of visits rather than documentation of impacts on the audience. He believes that this trend has been shifting over the past seven years as budgets for forestry Extension decrease and governments require more accountability for funds spent on Extension. Beck argues that accountability lies with administrators rather than practitioners, but he expects this to change in the next 5-10 years. Beck also perceives that, in Central Europe, as in many other parts of the world, there is a problem with the dual roles that forestry Extension staff are asked to play as both forest law regulator and educator and that this dichotomy makes evaluation difficult. Beck expects that the current emphasis on "quality management" that exists within the Bavarian Forest Service at all levels (internal evaluation) will eventually make outcome evaluation more acceptable by practitioners.

Evaluation at the Forest Research Extension Partnership in Western Canada consists of periodic corporate-level evaluations conducted by third party evaluators, needs assessment studies, and end-of-event surveys that measure perceptions of impact among participants, and tactical planning that includes evaluation criteria and measurements (FORREX 2004). Some Extension programs, such as the Master Tree Grower Program in Australia, have included evaluation as an on-going part of their programming (Reid and Stephens 2002). Generally, published documents refer to evaluations that are independent academic studies such as those on attitudes of Danish forest owners towards Extension services (Boon 2003), rather than being conducted internally as part of operations.

2.5 Definitions of program evaluation

Guba and Lincoln (1989) argue that there is no "right" way to define program evaluation. Patton (1997) describes three primary uses of program evaluation: to judge merit or worth of a program (such as audits), to improve programs (such as through adaptive management), or to generate knowledge. *Summative* evaluations are conducted at the end of programs to help decision makers decide a program's future or assess its impact. *Formative* evaluations are conducted during the life of a program to identify its strengths or weaknesses and improve its quality and effectiveness (Sanders 1994). Evaluations conducted to *generate knowledge* are usually conducted by academics examining trends or causal links across programs (Patton 1997).

Process evaluation, in contrast to program evaluation, is aimed at understanding the internal dynamics of program operations, not the outcomes (Patton 1987). Process evaluation addresses three major questions: what is the program intended to be, what is delivered in reality, and why are there gaps between program plans and program delivery (Scheirer 1994).

It verifies what the program is and whether or not it delivered as intended to the targeted recipients. Program evaluation, on the other hand, focuses on the "difference programs make." There is considerable confusion between process and program evaluation, with the former often being favoured because it is relatively easy to conduct and document. However, program evaluation provides the most useful information on the effectiveness and value of a program.

Program evaluation is related to policy evaluation, however programs typically include human resources and infrastructure while policies are more likely to be regulations or standards with or without infrastructure. Many policy evaluations are synonymous with process evaluations. The term "audit" is sometimes erroneously used synonymously with evaluation, however audits generally refer to financial assessments conducted to ensure that a standard set of accounting norms is followed. An exception occurs with forest management certification, where audits refer to an assessment of the practices of a manager against an agreed set of management standards, Generally, skills required for program evaluation are considerably broader than financial system appraisals.

2.6 Program evaluation competency

Several academics have investigated core competency requirements for Extension personnel, and most have discussed competencies related to program evaluation in particular (Beeman et al. 1979, Ritsos and Miller 1985, Summerhill and Taylor 1986). Most studies are limited to one state, however, and no study has targeted Extension practitioners within a single discipline such as natural resources across all states. Some categorize evaluation competencies with research competencies, while others categorize evaluation competencies with planning competencies, making it difficult to analyse evaluation as a stand-alone competency. Beeman et al. (1979) found that both county and campus-based Extension personnel in Florida rated program evaluation among the "least important" core competency categories for themselves. Another study (Ritsos and Miller, 1985) showed that competencies in research and evaluation received the lowest priority rating by urban Extension practitioners in Ohio. In a more recent study, however, Cooper and Graham (2001) studied perceptions of Extension practitioners and their supervisors in Arkansas and found that perceptions regarding competencies for evaluation received a "very important" rating for most county-based practitioners. This suggests that a major change may have occurred in the way that Extension practitioners view evaluation activities. However, it is difficult to draw conclusions from these studies as they might relate to NREPs across the U.S.. Some studies have examined competencies across disciplines within one state, while others categorize Extension practitioners with other county employees. In still others, evaluation is not separated from planning or research.

Several state Extension offices, such as Michigan, have developed core competency guidelines for their personnel relating to evaluation and other skills deemed essential to Extension work (Levine 2001). The evaluation competencies vary among three categories of

employees: "introductory career stage, early career stage and continuing career stage" employees. In other words, the longer the employee's length of service, the more advanced the levels they are expected to demonstrate in evaluation and other skills. Michigan State University Extension (MSUE) established a "Core Competency Assessment Program," a computerized, self-administered professional development tool for MSE Extension practitioners in 1994 (Levine 2001). The tool helps staff members identify their own strengths and weaknesses in program evaluation and other areas. MSUE also developed an on-line library designed to assist staff members in increasing their competency levels.

Radhakrishna and Martin (1999) examined training needs among Extension agents regarding program evaluation and research methodology. The study cited previous work that described barriers affecting frequency and level of program evaluation conducted by Extension agents. Barriers included lack of time, lack of resources, and limited expertise in evaluation methodology. The study showed that a substantial percentage of respondents want more training in program evaluation (Radhikrishna and Martin 1999). More than half indicated a moderate to very high need for training in topics such as developing evaluation plans and designing questionnaires. Only nine percent of survey respondents were in the natural resources discipline.

A study conducted by the Florida State University Extension Service in 1986 examined acceptance and adoption levels of program evaluation among Extension agents in Florida (Summerhill and Taylor 1986). Fifty percent of survey respondents indicated that they had done a greater number of evaluations in the most recent two years as compared to previous years, but respondents were not asked why. The percentage that reported using program evaluation results varied among number of years employed by the university (the more years, the more likely to have conducted evaluation in the previous two years). The study showed that participation in program evaluation training greatly enhanced the likelihood of Extension agents to conduct evaluation. Sixty percent of respondents that reported using formal evaluations in the previous two years did so to assess changes in knowledge, opinions (attitudes), skills, and aspirations (KOSAs, also called KASAs) in the target audience, while 35% assessed higher-level impacts such as changes in behaviour in the target audience. The report does not indicate whether natural resource Extension agents were included in the survey. The study also does not reveal the nature of training received, nor the details of evaluations conducted.

A study conducted by Mohamed (1998) to determine perceptions of Extension agents in Ohio about performance appraisals showed that program evaluation and impact documentation criteria received the lowest mean score for current and future use among performance appraisal criteria. Jha (2001) concluded that Extension educators do not like to do evaluation, have inadequate training to do so, perceive it as something separate from programming, and face time constraints.

Douglah et al. (2003) conceptualized evaluation competency as an overlapping set of evaluation skills, knowledge, and positive attitude toward evaluation (i.e. motivation and commitment to use evaluation for self-improvement and accountability). They describe how the quality of evaluation decisions and actions can be the result of an individual's ability to integrate these components. The survey conducted by Douglah et al. (2003) of 354 Extension county-based faculty and staff in Wisconsin showed that 36% of respondents cited lack of time as the reason why more evaluation was not being conducted. Twenty percent said that they would perform more evaluations if they believed it was valued by supervisors or felt it influenced program decisions. Twenty percent said their reason for not conducting more evaluation was because they doubted its value. Seventeen percent cited lack of tools such as templates, and 16 percent cited lack of knowledge, skills, or confidence.

Although limited in scope, these studies provide insight regarding the potential influence of training, incentives, and other factors in increasing level of program evaluations among Extension professionals. These studies show that both characteristics of the individual (such as length of years in the profession) and characteristics of the working environment are associated with the level of evaluation conducted.

2.7 Evaluation mainstreaming and related concepts

Several authors have recently described the integration of evaluation into everyday operations of organizations as shown in Table 2-3.

Table 2-3 Evaluation concepts related to mainstreaming

Evaluation Concept	Authors/organization
Mainstreaming	Barnette and Sanders 2003
Evaluation Capacity Building (ECB)	W.B. Compton et al. 2002, World Bank 2002
Internal evaluation	Sonnichson 2000
Evaluative Inquiry or Learning Organizations	Torres and Preskill 2001, Senge 1990
Enculturation	Toulemonde 1999
New Public Management	Ferlie et al. 1996

Barnette and Sanders (2003) defined mainstreaming as the "process of making evaluation an integral part of an organization's everyday operations." Their definition includes the integration of evaluation in work ethics, culture, and job responsibilities at all levels of an organization. Barnette and Sanders proposed several indicators of successful mainstreaming in an organization:

- when the Chief Executive Officer of an organization distributes a list of organizational values that includes evaluation;
- when orientation for new employees includes information on their role in evaluation;

- when consumers of programs ask for evaluation data; and
- when evaluation appears on "every agenda".

Barnette and Sanders (2003) explored barriers to mainstreaming evaluation, citing a lack of leadership, lack of consultants who focus on evaluation capacity-building, lack of utilization of evaluation results in decision making, lack of acceptance of the cost of evaluation as a normal cost of doing business, and lack of awareness of benefits. They recommend identifying and working with allies such as foundations, professional evaluation associations, businesses, and education, government, and human services organizations. They recommend compiling and publishing case studies of organizations that have successfully mainstreamed evaluation to allow others to learn from their experience, and supporting and recognizing organizations that are trying to mainstream evaluation into their organizations. The 2001 American Evaluation Association conference theme was "Mainstreaming Evaluation" and resulted in several influential papers, compiled by Barnette and Sanders (2003).

Willliams and Hawkes (2003) describe the task of leadership in mainstreaming as a balancing act between requiring and inspiring employees to conduct and use evaluation. They refer to capitalizing on existing access points such as encouraging more donor organizations to make evaluation a part of their funding requirements as well as including evaluation experience as a hiring criterion. Williams and Hawkes (2003) also recommend more user-friendly evaluation reports, evaluation certification as a way of helping people in the mainstream recognize and accept evaluation, and building a common language. Overall, the authors agree that evaluators need to be more customer-focused in order for evaluation to become part of organizational culture. Sonnichsen (2000) believes that when evaluation activities are legitimized as routine and results are used in decision-making, mainstreaming has been accomplished. The process of mainstreaming is not a linear process, but a staged, iterative process that does not happen overnight. Evaluation capacity building, defined as the intentional work to continuously create and sustain overall organizational processes that make quality evaluation and its use routine, is discussed separately in the next section (Compton et al. 2002).

Sonnichsen (2000) defined "internal evaluation" as the process of reviewing administrative, operational and environmental activities by employees who are trained and mandated to gather independent empirical information to aid in organizational improvements." He states that three factors have led to an increased need for internal evaluation capacity: the demand for improved performance and service delivery of organizations, the availability of information technology for processing large volumes of data, and the increased complexity of organizations. He believes that with senior-level management commitment to continuous examination of operations, internal evaluators (those who work inside an organization) can provide important information crucial for high-level decision-making. Internal evaluators can serve as catalysts for organizational learning in ways that external evaluators who are not linked to strategic planning and budgeting cannot. To be successful, the internal evaluation

function must be fully integrated with other critical organizational functions. Evaluators must be more than data collectors and analysers; they must also be advocates of the application of evaluation findings. Sceptics question internal evaluators' abilities to be independent, objective, or adequately scientifically rigorous, but advocates see internal evaluation as more useful and relevant to decision making, while still possessing the ability to have adequate independence and relevance if their autonomy is protected by top-level administrators and if they are positioned appropriately in the organization (Sonnichson 2000). The practice of internal evaluation can serve as the basis of organizational learning, enhancing an organization's ability to continually expand its capacity to create its future (Senge 1990).

A body of literature has developed in the past twenty years that reflects the assumption that cultivating a workplace that is committed to learning and which engages in reflective practice leads to more productivity and higher level of accomplishment of desired goals. Torres and Preskill (2001) assert that the "evaluative inquiry" approach shares several attributes with Barnette and Sanders' mainstreaming approach but it focuses specifically on structures within organizations relating to their ability to incorporate evaluation as a practice. Evaluative inquiry is described as a continual pro-active process of internal investigation regarding organizational processes that is routinely conducted by the organizations' members. Torres and Preskill (2001) contrast this approach with more traditional evaluation methods that involve independent, third-party audits of organizational functions. In evaluative inquiry, leaders focus on developing the ability, interest, and involvement of organizational members to critically examine the organization's processes using evaluation methodology.

Torres and Preskill (2001) argue that an organization's communication systems, leadership approach, team structure, reporting structure, and willingness to reward risk-taking provide the foundation for evaluative inquiry. They believe that support for evaluative inquiry must come from the top of the organization as well as mid-level leaders. How information is communicated down the line is a key determinant of the extent an organization is likely to succeed with evaluative inquiry.

Several authors have linked development of evaluation capacity and culture with the cultivation of a "learning organization" (Douglah et al. 2003, Senge 1990). Others describe how evaluation in public agencies in the 1960s and 1970s emphasized formal "top-down" external evaluation procedures for the purposes of accountability, the results of which were largely unused by organizations for self-improvement (Torres and Preskill 2001). Douglah et al. (2003) describe how more recent participatory evaluation approaches (involving all levels of an organization) over a sustained time period can be empowering, useful, applicable, and lead to greater productivity and individual growth. Douglah et al.'s (2003) model for the development of a learning culture includes a) cultivation of evaluation skills, b) policies that promote evaluation, and c) the presence of a shared vision of evaluation that is accepted by all members of the organization. Douglah et al. (2003) believe that previous attempts to increase skills have been insufficient. In addition, they believe that:

- individuals and groups within organizations must believe that evaluation is valued by the organization by observing that the organization takes action on its verbal commitment to evaluation;
- regular dialogue about evaluation within the organization is encouraged, not just to communicate, but to examine topics in depth; and
- leaders are willing to cherish both successes and failures as learning opportunities.

Douglah et al. (2003) assert that not all members of an organization need to be evaluation experts, and that a bell-shaped curve, with percent of employees on the vertical axis and evaluation knowledge, skills, and attitudes on the horizontal axis, illustrates an ideal evaluation competency within an organization. A small number of doubters and proctors lie at the far left of the curve (those who believe evaluation is a waste of time and those who would cooperate with evaluation but not initiate it themselves), and a small cadre of specialists/consultants and scholars lies on the far right hand side of the bell curve. In the middle, the largest group is considered evaluation practitioners. Douglah et al. (2003) recommend that organizations strive to decrease the percentage of doubters and proctors and increase the proportion of staff members who are evaluation practitioners, specialists, and scholars. Ideally, operational staff members serve as the evaluation practitioners and they use a mixture of evaluation methods to assess their own programs, have positive attitudes towards the role of evaluation in their work, and believe that their organization positively rewards them for conducting evaluation.

Toulemonde (1999) studied ways that administrators create incentives and constraints to encourage employees to conduct program evaluation. Through a series of case studies, he concluded that a combination of "sticks" (constraints) "carrots" (incentives) and "enculturation" (making evaluation part of the work culture) offered the most effective formula for inducing evaluation among practitioners. According to Toulemonde, the combination of requiring practitioners to establish an evaluation work plan and conducting a systematic quality assessment (sticks), and earmarking funds for evaluation, granting local control for evaluation data, conducting systematic involvement of stakeholders, and decentralizing the evaluation process (carrots) provide a crucial blend of incentives for inducing evaluation. Toulemonde found that increasing practitioner knowledge about evaluation alone had no significant impact on the likelihood of staff members to conduct evaluation. Compulsory evaluation was ineffective where there was a lack of skilled professionals to assist practitioners in conducting evaluations.

Toulemonde also considers "building an evaluation culture" within an organization as a third important factor affecting the likelihood of practitioners to conduct evaluation. He believes that deeply rooting evaluation in administrative values and norms leads practitioners to overcome reluctance even when the evaluation results are likely to contradict their self-interest. Toulemonde suggests that the success of establishing an evaluation culture depends

on the quality of communications about evaluations; he suggests communicating through workshops, newsletters, journals, training courses, demonstration, projects, prizes, and awards. As a result of the enculturation process, people feel that they belong to a community of people that trusts that evaluation is part of sound management. He argues that clever administrators use both carrots and sticks, but acknowledges that only enculturation can overcome resistance to evaluation. Toulemonde does not detail a formula for enculturation, mainstreaming, or evaluation capacity-building but leaves the discussion to later authors such as Barnette and Sanders (2003), Duigan (2003), and Compton et al. (2002).

There is trend towards the restructuring of some public agencies, notably in the United Kingdom and U.S., as well as recently in B.C. (such as with the transfer of health service responsibility to regional health authorities) into what can be considered more private-sector ideology (Ferlie and Hartley 2003, McDavid 2001). Labelled "*The New Public Management*" (NPM), it is an emerging organizational phenomenon in the literature. Advocates argue that NPM leads to a new brand of market-based accountability where consumers of government programs and services have greater choices among suppliers and producers, and level of demand for products and services becomes the indicator of success. Critics suggest that accountability is more difficult under a system influenced by private sector approaches because publicly produced products and services such as education and programs for low-income seniors are often linked social values that work outside the rational approaches of supply and demand (Ferlie et al. 1996). Nonetheless, structural reorganization under NPM is accompanied by a need to establish new indicators of success, new mechanism to measure indicators, and new players to be accountable for measuring impacts. Ferlie et al. (1996) assert that public management reform is a global phenomenon, and the principles have been argued as being relevant to public-sector research in countries such as Switzerland and Germany (Schedler 1995). The concept has found favour with groups such as the Advisory Council of the Swiss Institutes of Technology, and attempts were made to introduce it into forestry research and extension in Switzerland in the 1990s (J. Innes, Professor of Forestry, University of British Columbia, B.C. pers. comm., July 2004)

2.8 Evaluation Capacity Building (ECB)

The concept of evaluation capacity building (ECB) is favoured by some scholars as an emerging field of practice because of its link to organizational development. Compton et al. (2002) refers to the concept as the intentional work to continuously create and sustain overall organizational processes that make quality evaluation and its use routine. While relatively little has been published that uses ECB terminology and that focuses on ECB as a distinct field from program evaluation, several authors (e.g., Mckay 2002, King 2002, Compton et al. 2002) urge the acceptance of ECB as a legitimate focus within the evaluation community because it requires a unique set of skills, tasks, and responsibilities and behaviours. ECB is most often linked to social and economic development, and the concept is particularly used by

international development organizations. Barnette and Sanders (2003) believed that it is possible – and desirable - to build evaluation capacity without making evaluation part of an explicit business within an organization; it simply becomes integrated into planning and programming as a way of doing business. Duigan (2003) believes the term mainstreaming means "trying to put evaluation in center stage" or possibly an attempt by evaluators to raise the profile of evaluation as a profession. He believes that 'giving evaluation away' by sharing approaches without necessarily being called evaluation by those who use them is the essence of ECB. In other words, the more profile that evaluation has within an organization, the less it is fully mainstreamed.

Duigan (2003) suggests that building evaluation capacity depends on a) selecting an appropriate evaluation model for the organization, b) developing skills at every level in the organization, and c) identifying strategic high-priority evaluation questions by the highest levels within an organization. He suggests that selecting among the many evaluation approaches, such as goals-free (as described by Scriven 1991), utilization-focused (described by Patton 1997), and empowerment evaluation (described by Fetterman 2000), that best fit the culture of the organization is an important first step in building evaluation capacity. Duigan asserts that training sessions and manuals are key to developing staff evaluation skills; he provides examples ranging from brief evaluation presentations for staff members to week-long workshops for managers and service providers. Finally, he believes that too often leaders state that "every program needs an evaluation" without determining which program is highest priority and for what purpose. He recommends expending resources on the highest priorities within an organization.

A world leader in international development and development financing, the World Bank has a long history of developing evaluation capacity among public institutions in borrower countries as an integral part of its development agenda (World Bank 2002, Compton et al. 2002). Since 1994, the World Bank has been developing and implementing a strategy for evaluation capacity building with the goal of helping countries build and use internal monitoring and evaluation (M & E) systems. The World Bank aims to help countries use M & E information to inform government decisions, enhance transparency and accountability of government investments, and ultimately maximize government performance. The support includes evaluation training, assistance in developing performance indicators, assisting in developing and improving national statistical systems, and financial management and auditing systems, as well as helping encourage a demand for information generated through M & E activities. The World Bank has a number of publications to guide development of M & E capacity, such as checklists and toolboxes.

The Canadian International Development and Research Center (IDRC), headquartered in Ottawa, provides a model of a research and development organization that has developed internal capacity and infrastructure for evaluation (Horton et al. 2003). IDRC is a public corporation created by the Canadian Parliament in 1970 to provide assistance to

developing countries to address social, economic, and environmental problems. IDRC's Evaluation Unit was started in 1992 and currently has seven evaluation staff members who promote evaluation as a planning and management tool, help build evaluation capacity among its country partners, and assess the impacts of research for development. Evaluation at IDRC is considered as "decentralized" with individual program and project managers holding responsibility for evaluation while the Evaluation Unit staff members provide coordination and support. All IDRC programs are expected to have evaluation plans and the program staff is expected to include evaluation results in reporting. IDRC's evaluation system involves guidelines indicating who has responsibilities for what kind of monitoring, and defines purposes of various types of evaluation. The Evaluation Unit provides guiding principles as well as evaluation tools for program and project staff. The guiding principles state that evaluations should leave an increased capacity to use evaluation findings and should enlist the participation of relevant stakeholders. The IDRC evaluation strategies include working in partnership with other donor organizations to build the evaluation capacity of its developing country partners.

Milstein and Cotton (2000) identify five basic elements of ECB within an organization:

- motivation forces (provided by the administration to motivate employees)
- organizational environment (structures within an organization that lend themselves to evaluation)
- workforce and professional development (courses, workshops, and mentoring)
- resources and supports (written material websites available to help staff with evaluation); and
- and reporting on lessons learned from experience (communication about evaluation successes and challenges).

2.8.1 Case studies of Evaluation Capacity Building

1) United Way of Toronto. One example of an evaluation capacity-building initiative is a five-year project led by the United Way organization of Greater Toronto, Canada (2000-2004) to improve the community-based social services provided by non-profit agencies funded by the United Way (United Way 2003). Called the Program Effectiveness Organizational Development (PEOD) project, it has included free evaluation workshops, short-term evaluation consulting upon request, distribution of written evaluation resources, and a web-based clearinghouse for 250 non-profit community-based agencies in Toronto. The program began with a survey of agency leaders to determine the level of evaluation skills and activities. The survey identified gaps in capacity ("outcome measurement," "setting indicators," and "using appropriate measurement tools"). Based on the results of the survey, United Way of Toronto developed their own desired outcomes, such as:

- agencies increase knowledge and skills in outcome measurement and use of logic models;

-agencies identify enablers and constraints to implementation of evaluation; and
-agencies integrate program effectiveness processes into management and strategic development throughout the organization.

Indicators included the percentage of members demonstrating new knowledge, the percentage of agencies developing logic models, and the percentage of agencies reporting integration of program effectiveness processes into management and strategic development throughout the organization. Through site visits, interviews, a survey mailed to agencies one year after attending the first workshops, and by analysing funding submissions from the agencies, United Way determined that the initiative was successful in raising the level of knowledge and skills in evaluation and changed the behaviour of organizations. The survey showed that 81% of agencies exposed to the program had taken steps to establish an organizational structure to coordinate outcome measurement work, including creating a committee or assigning the task to an existing committee or staff and 86% had drafted or were in the process of drafting a program logic model (evaluation tool).

United Way's assessment of their initiative also identified challenges expressed by the agencies in integrating evaluation into their regular routines, including:

- lack of mechanisms for stimulating "buy-in" by staff;
- lack of skills and time to develop organizational structures, processes, and policies for evaluation;
- minimal administrative infrastructure to collect, process and analyse evaluation data; and
- service delivery models that limited their ability to capture data on or from transient or anonymous clientele.

United Way of Toronto concluded that integration of evaluation into organizational structures and cultures of non-profit agencies requires considerable time, effort, guidance from leadership, access to technical support, and patience to move from awareness to behavioural change but concluded that the investment in evaluation capacity building by funders yields better service outcomes.

2) Centers for Disease Control and Prevention. Milstein et al. (2002) describe evaluation capacity building of the Centers for Disease Control and Prevention (CDC) headquartered in Atlanta, Georgia. The CDC has a $2.8 billion budget and is funded by a wide variety of public and private granting partners. It employs over 8,000 people in 13 centers in an effort to improve public health and prevent disease worldwide. Coincident with the passage of the U.S. Government Performance and Results Act (GPRA 1993), CDC leaders attempted to enhance the organization's capacity in evaluation over a four year period between 1997 and 2000 (Milstein et al. 2002). Requirements of the GPRA as well as changes in health indicators (i.e. a recognition that health is linked to larger community and environmental conditions and

systems, not just individual behaviours and conditions) led to increased attention being paid to strategic planning and evaluation as well as a new awareness of evaluation capacity gaps throughout all levels of the organization. Through a comprehensive assessment involving listening sessions, workshops, interviews, the CDC concluded that it needed to:

- develop a working definition of evaluation within the organization and health sector;
- formulate an evaluation framework;
- create a focal point for leadership on evaluation within the organization;
- build an evaluation-literate workforce;
- nurture a culture of accountability; and
- share evaluation information across the agency.

The assessment revealed that evaluation was being done with significant variability across units in the organization, a reflection of the wide diversity of partners, funding arrangements, and health topics. Evaluations were often conducted by subcontractors independently of program staff and leaders. The CDC established an Evaluation Working Group (www.cdc.gov/eval) to solicit recommendations from staff and partners and develop plans for building evaluation capacity. The working group, comprised of representatives from every center, institute or office and open to all interested parties, established an evaluation framework and continues to provide evaluation leadership and consultation to program staff and stakeholders (CDC 2004).

The first tasks of the working group included setting up an evaluation website to serve as a focal point, listening to input through extensive workshops and interviews, and conducting a literature review of organizational evaluation frameworks. After six months to compile recommendations and ideas, a national satellite training program was used to introduce the draft framework to about 10,000 staff members and partners. The working group logged over 200 observations and recommendations and sorted them. The input led to a set of 12 recommendations relating to evaluation practice and institutional changes needed to enhance evaluation. The recommendations were:

- designate coordinators and staff dedicated to evaluation;
- dedicate funds to evaluation;
- create logic models that link work of all levels of the organization;
- create technical assistance directories for staff assistance with evaluation;
- create incentives systems for evaluation practices;
- designate an organizational lead or champion for evaluation;
- develop an evaluation consultation corps for staff assistance;
- train staff in evaluation methodology to establish a common vocabulary and understanding of evaluation;
- produce materials to support practice and share findings with stakeholders; and
- sustain leadership over time, not just in the early months and years of the initiative.

The authors cite the CDC's transparent, neutral process for the broad support it received across the organization for the evaluation framework. The "listening sessions" (such as brown bag sessions, group interviews) and meetings involving whole branches or divisions were cited for their role in the success.

Recurring themes arising from the process were:

- a need for a common evaluation language;
- a need for clearer communications of program goals and expectations from the top down;
- reframing the organizational culture to that of a learning organization where failures are an expected and accepted as part of management, so evaluation is not resisted but seen as a part of sound management; and
- capacity building efforts cannot proceed in isolation of other forces affecting change in the organization, and that evaluation strategies should align with those changes (changes such as increasing community involvement in health activities and improved information and data management systems).

The working group operated under several key principles: 1) evaluation is best when considered at the program level and integrated in program operations; 2) agency-wide coordination is essential; 3) adequate resources are needed to conduct sound evaluations; 4) greater capacity is needed for internal evaluation at all organizational levels; 5) support is needed for partners to participate; and 6) leadership, incentives, and long-term culture change are necessary.

Field testing of the framework was done in 1998-1999 and interest throughout the organization was high. The framework was revised using consultations with staff and partners. The final framework was disseminated in over 60,000 copies worldwide by email or mail. Numerous adaptations have been published by various units of CDC and its partners using the framework as a guide. Other spin-offs such as revisions to staff training and orientation have utilized the new framework.

The CDC is now seen as an important source of information and expertise in program evaluation by other institutions. The Evaluation Working Group remains a viable force within the CDC, with a continuing mandate to provide products and services related to improvement of evaluation and institutional support for evaluation. The working group affirmed that the process of evaluation capacity building is a permanent and on-going task, not a one-time initiative. The CDC received the President's Prize at the annual meeting of the American Evaluation Association in 1999 in recognition of the framework and the process of developing it.

CDC organizational leaders learned some important lessons that provide insight to other organizations seeking to improve their evaluation capacity: 1) people and organizations

learn about evaluation by doing it. Taking time to reflect on lessons by comparing against program evaluation standards such as the Joint Committee on Standards for Education Evaluation is also a key to effective capacity building. 2) Resistance to evaluation can be overcome if leaders can circumvent the natural instinct of their staff to avoid criticism by emphasizing the learning aspect of evaluation, by depersonalizing results, and sharing results of evaluations of their own work even when it may report negative findings. 3) Building evaluation capacity change is incremental and gradual, and leaders should take advantage of "teachable moments" such as reorganizations to introduce changes. 4) Knowing the organization's culture is one of the most valuable assets in developing an evaluation culture.

Further, the case study showed that evaluations' special role is to draw together the processes that are usually fragmented in organizations such as planning, research, infrastructure building, policy setting, marketing, decision making and leadership. All these systems are necessary in the process of building capacity for evaluation within organizations.

2.9 Bennett's Hierarchy

In his report, "*Analyzing Impact of Extension Programs*," national CSREES Extension leader Claude Bennett introduced an evaluation model that has become a standard tool for planning and assessing impacts of Extension programs (Bennett 1976). "Bennett's Hierarchy of Evidence" describes a series of staircase levels of evidence of program impacts, beginning at the bottom step with "inputs" (allocation of resources to a program) and progressing to the top step, the "end result" (measuring impacts of a program on long-term goals or conditions). Evidence of program impact at each ascending step is progressively more substantial albeit more difficult, costly, and time-consuming to measure. According to Bennett, higher-level evaluations provide stronger evidence of impact than "lower level" evaluations.

At the lowest level on the Hierarchy, an Extension practitioner measures and reports on the amount of dollars allocated to a project as an indicator of program success (level 1: *Inputs*, as shown in Figure 2-2). While these data are easy to obtain, they do not say much about "what difference" the program makes. Higher in the Hierarchy, however, Extension practitioners measure changes in knowledge, opinions (also called attitude), skills and aspirations of the target audience as a result of their program (level 5, "*KASA*"). This kind of evidence shows that the program leads to more substantive change than lower points on the Hierarchy.

The highest level of Bennett's Hierarchy shows process toward a long-term objective such as increased species diversity in areas where species loss has occurred (level 7: *End Results*). In most cases, the attainment of Level 7 is only theoretical since it is usually impossible to isolate other factors that may influence the end result (for example, changes in regulations could also lead to increased number of species).

Bennett also asserted that the "harder" the data available at each step, the more an evaluation can be relied upon for decision making. Examples of hard data include direct

observations of changed behaviour and results of "pre" and "post" skills tests. "Soft data" by contrast, include information such as retrospective reports by farmers regarding their changes in practice or perceptions of skills gained. Furthermore, evaluations using controlled experiments provide stronger evidence than naturalistic studies, or in other words, surveys provide weaker evidence than methods used in experimental research.

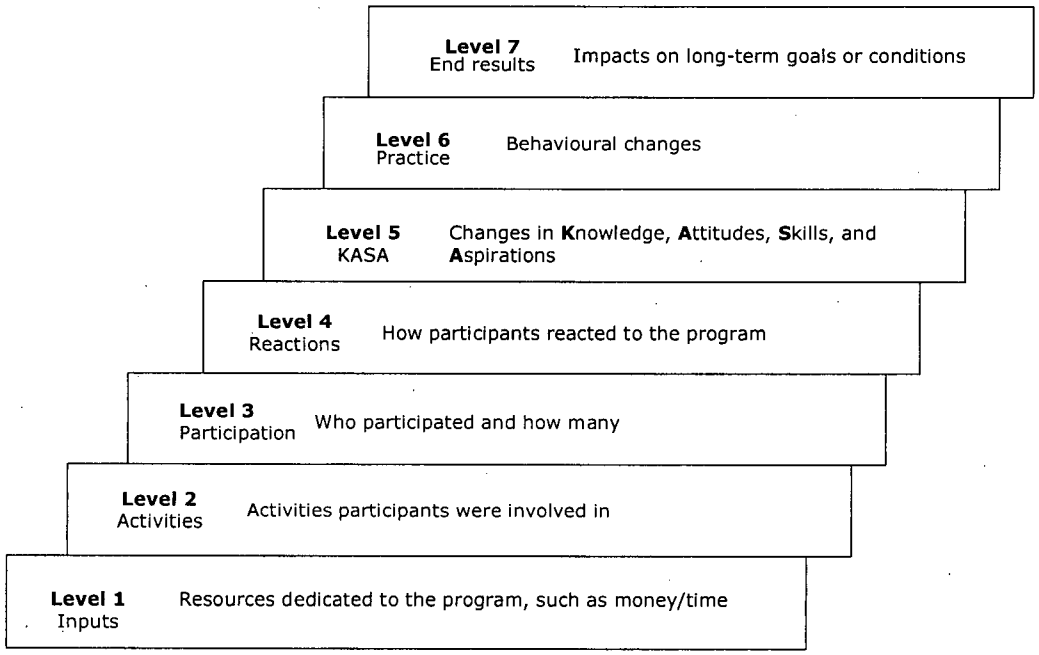

Figure 2-2 Bennett's Hierarchy of Extension Program Evaluation. Source: Bennett, Claude F. 1977. Analyzing Impacts of Extension Programs. Washington D.C. U.S. Department of Agriculture Extension Service.

While Bennett's Hierarchy is still currently widely used by Extension organizations for evaluation, some scholars, including Bennett himself, have modified the Hierarchy to reflect the increasing complexity of extension program goals (such as achievement of sustainable agriculture) (Suvedi et al. 2003)) and recognition of the links between program planning and evaluation. Bennett and Rockwell (1995) developed and tested a modified version of the Hierarchy called the "Targeting Outcomes of Programs" (TOP) model that is used by some Extension organizations to integrate program planning and evaluation (Bennett and Rockwell 2004). The TOP model is similar to Bennett's Hierarchy except that it includes an additional component to show the how the seven levels are also appropriate in program planning. The right-hand side of TOP's "V" shape is the original Hierarchy; the addition of a left-hand side of the "V" depicts the seven levels in reverse order to show levels of planning. In other words, the levels of program planning are shown by the seven levels in descending order on the left-

hand side of the "V," while evaluation is depicted by ascending the model on the right-hand side of the "V."

Suvedi et al. (2003) integrated Bennett's Hierarchy with a model often used by international development organizations in evaluating interventions directed at sustainable development, called the "Driving Force-State-Response" Model. They proposed this hybrid framework for assessing the two highest levels of the Hierarchy (practice changes and impacts on long-term goals - levels 6 and 7) for extension programs such as those aimed at complex constructs such as sustainable agriculture, for which there are social, environmental, and economic indicators of success. Examples of driving force, state, and response indicators are policy changes, increased networking levels among farmers, and increased investments in research and education; these are indirect forces that influence longer-term outcomes.

Despite these modified versions of Bennett's Hierarchy, the original model remains a common tool among Extension organizations for evaluation (Bennett and Rockwell 2004).

2.10 Theories of motivation

Methods for studying NREP program evaluation behaviour among NREPs are grounded in several motivation theories that developed in the fields of psychology in the early and mid-20th century. Some of these have been used for several decades as the basis for employee incentive programs in corporations throughout North America. Motivation has been defined as "the complex of forces, drives, needs, tensions, states or other internal psychological mechanisms that start and maintain activity toward the achievement of personal goals" (Hoy and Miskel 1978). Luthans (1995) defined motivation as "a process that starts with a physiological or psychological deficiency or need that activates behaviour or a drive that is aimed at a goal or incentive." Work motivation is considered a "set of energetic forces that originate both within as well as beyond an individual's being to initiate work-related behaviour, and to determine its form, duration, and intensity" (Pinder 1984). Katzell and Thompson (1990) define work motivation as "a broad construct pertaining to the conditions and processes that account for the arousal, direction, magnitude, and maintenance of effort in a person's job." For this thesis, I will assume that work motivation is an interactive set of forces derived both externally and internally that influence work behaviours.

Motivation theories are widely used in the education and organizational development fields to help educators and managers understand what leads students and employees to certain behaviours (Lawton and Rose 1994; Katzell and Thompson 1990). Managers desire to know what factors motivate employees to exhibit certain behaviours so that they can create a work environment that leads to higher work productivity (Pardee 1990). The purpose of motivation theory is to explain how behaviour is initiated, sustained, directed, and stopped. It is assumed that a highly motivated individual will perform better than one with low levels of motivation.

There is no single unifying theory of human motivation in the literature and the divergent theories represent varying worldviews (O'Neill and Drillings 1994). Contrasting theories provide the basis for debate between "work environment" or an "individual's characteristics" as the explanatory factors behind work motivation and behaviour. Three categories of motivational theories considered in this study include 1) non-cognitive theories such as reinforcement theory (Skinner 1953 and Pavlov 1960) and life-course theory (Featherman 1983); 2) cognitive theories such as personal investment theory (Maehr and Braskamp 1986), goal theory (Sims and Lorenzi 1992) and self-efficacy theory (Bandura 1977); and 3) hygiene theory (Herzberg 1991) (see Table 2-4). Cognitive theories emphasize conscious goal setting, while non-cognitive theories emphasize stimulus/response factors. Hygiene theory relates to factors in a work environment that influence job satisfaction and dissatisfaction. Motivational factors have been categorized as either *exogeneous* (focused on variables that can be manipulated by external agents) or *endogenous* (variables such as attitude and other personal attributes inherent in the individual or that change indirectly in response to external variables) (Katzell and Thompson 1990).

Table 2-4 Theories related to work motivation

Theory category	**Theory name**	**Author/Researcher associated with theory**
Non-cognitive	Reinforcement	Skinner 1953, Pavlov 1960
	Life-course theory	Elder 1985
Cognitive theory	Personal investment theory	Maeher and Braskamp 1986
	Goal theory	Sims and Lorenzi 1992
	Self-efficacy theory	Bandura 1977
Hygiene theory		Herzberg 1991

2.10.1 Non-cognitive theories: reinforcement theory and life-course theory

Sometimes called operant or learning theory, the reinforcement theory is based on the work of Pavlov (1960), who developed the theory of conditioned reflexes in the late 19th century and Skinner (1953), whose theories about the influential nature of consequences on behaviour dominated the psychology literature for many years (Locke and Latham 1994). Reinforcement theories assert that behaviours are controlled by the existence of stimuli that induce a learned response in an individual. Actions taken by an individual in a given situation are influenced by the consequences of behaviours from the same or similar stimuli in the past (e.g., "I was rewarded for a given behaviour in the past, so I will do it again"). Managers apply reinforcement strategies (rewards and disincentives) to increase or maintain desirable behaviours and decrease or eliminate undesirable behaviours. This framework began to enter organizational literature and teaching texts in the 1970s, and was especially focused on manager-employee relationships. It attempted to explain how managers influence behaviour of employees by structuring rewards (also called incentives) and disincentives and has formed

the basis of many business leadership programs (Sims and Lorenzi 1992). Behaviourist theories underlie many common incentive programs such as sales commissions and employees of the month recognition programs (Kohn 1993). Behaviourist practices are still widely used in the work environment, although there is belief by some that these practices do not lead to true motivation (Kohn 1993, Locke and Lantham 1994).

Sims and Lorenzi (1992) believe that rewards and disincentives are used by organizations in five categories: material, symbolic, social, task, and self-administered. Material incentives constitute a direct or indirect financial reward to the employee. Symbolic refers to psychological rewards and can be either tangible or intangible. Social rewards are interpersonal. Task rewards include those directly related to the design of job tasks, such as levels of autonomy or scheduling flexibility. Examples of the categories are in Table 2-5.

Table 2-5 Examples of organizational rewards and disincentives

Rewards	**Examples**
Material	Wages
Symbolic	Plaques, awards, promotions
Social	Praise, feedback, recognition, positive verbal and non-verbal communications
Task	Enriched job, preferred task assignments, new job responsibilities, flexible work hours
Self-Administered	Self recognition, self-praise, sense of accomplishment
Disincentives	
Material	Dismissal from job, loss of opportunity for promotion/tenure, not rehired for future programs
Symbolic	Involuntary transfer to another location
Social	Non-verbal disapproval, social isolation, lack of feedback
Task	Boring, repetitive work, close supervision or control
Self-Administered	Sense of failure, guilt, shame, sense of "letting the group down"

(adapted from Sims and Lorenzi 1992)

The assumptions underlying the behaviourist approach to employee motivation are debated in the literature (Locke and Lantham 1994, Kohn 1993, Brody 1983). Kohn (1993) asserts that while the majority of U.S. corporations use incentives to motivate employees, a growing collection of evidence shows that incentives do not motivate employees to perform their work, they only motivate employees to receive the reward (Kohn 1993). She asserts that behaviourism is fundamentally flawed because the assumptions behind them are flawed. Critics believe that while incentives may modify behaviours in the short term, they do not lead to sustained changes in attitudes and behaviour over time. Incentives do not create commitment to an action. A meta-analysis of 98 studies about incentives (Guzzo et al. 1985) showed that goal setting had far greater impact on productivity than salary incentives. Kohn

believes that incentives are used because they are easy to implement and typically provide immediate results, but that they provide only temporary compliance, not motivation.

The life-course theory says that personal goals vary with age and stage of career (Elder 1985). Elder defines life course as the interconnected trajectories that a person has as he or she ages through life. He believes that each career stage and age has a distinct set of pressures that lead individuals to behave in certain ways. He also states that interventions are most effective if they are sensitive to the developmental needs and capabilities of particular age periods in the life span (Elder 1985). In the context of Extension evaluation, this theory would assert that that the variation in evaluation attitudes and behaviours of NREPs could be explained by a person's age and stage of life.

2.10.2 Cognitive theories: goal theory, personal investment theory, and self-efficacy theory

Cognitive theories assume that people make behavioural decisions through conscious (cognitive) intentions and assessing potential gains and losses of their choices of behaviour. Goal theory as described by Sims and Lorenzi (1992) and Katzell and Thompson (1990) is based on the premise that the process of developing clear, specific and challenging but attainable goals is a major motivational force in individuals. While reinforcement theorists believe that learned response is the key motivational factor, goal theorists claim that one's personal goals are the most critical precursor to behaviour. This theory says that an individual's motivation to perform is associated with their sense of freedom and expectation to set their own goals. In a work context, this theory asserts that a management style that relies on employee initiative in setting and meeting goals is more effective in motivating employees to perform than reinforcement. Employees with goals perform at higher levels than do people without goals. More challenging goals lead to higher performance than easier goals (Sims and Lorenzi 1992).

Extension administrators who apply the goal theory of motivation are more likely to work with employees to establish goals relating to their programs and evaluation efforts than to rely on rewards and disincentives to affect evaluation behaviour.

With personal investment theory (Maehr and Braskamp 1986), it is argued that motivation to perform a task depends on the value or meaning that an individual places on the task and that one's estimates of personal control over a situation, self-competence, and desired outcomes are determinants of behaviour. This theory considers both individual and contextual (environmental) factors.

In self-efficacy theory, work behaviour is determined by an individual's assessment of their ability to accomplish a task (Bandura 1977). In other words, the level of engagement in the task is associated with confidence in one's ability to meet outcomes. Self-efficacy is not a stagnant or "stand alone" concept; it is dynamic and influenced by the nature of the task, the level of desirability of the outcome, perceptions about access to resources, sense of control,

and past successes with the task. Unlike self-esteem, which is a more global concept, self-efficacy refers to confidence in accomplishing specific tasks.

Socialization theory suggests that individuals develop attitudes and patterns of behaviour through early socialization. Blackburn and Lawrence (1995) use this theory in their study on university faculty motivation, predicting that faculty attitudes towards conducting and publishing research are acquired during professional training. Socialization theory, therefore, would predict that NREPs who are educated in institutions where program evaluation is an explicit value and program emphasis will be more likely to value and use evaluation than those who attended institutions that did not emphasize this.

2.10.3 Herzberg's hygiene theory

Frederick Herzberg (1923-2000) devised a motivation framework during the 1950s and 1960s for factors that he believed affected work attitudes and behaviours that built on the cognitive and non-cognitive traditions. He found that certain factors directly related to their work environment tended to cause a worker to feel satisfied or dissatisfied with his or her job. Herzberg's premise was that certain factors such as company policy, supervision, interpersonal relations, working conditions, and salary are "hygiene factors" rather than motivators. The absence of these factors can lead to job dissatisfaction but their presence does not necessarily lead to job motivation. He offered six factors he considered motivators or "satisfiers": 1) achievement, 2) recognition, 3) the work itself, 4) responsibility, 5) advancement, and 6) growth, as described by Pardee (1990).

Herzberg attempted to distill what he considers as the myths of incentives and disincentives (Herzberg 1991). He distinguished between "motivation" and "movement," saying that both incentives and disincentives may lead to movement but not necessarily to motivation. He reviewed several common personnel practices that are designed by management to motivate employees and describes how these practices may lead to short-term movement but not sustained motivation. He uses an analogy of a battery-charged generator:

> *"I can charge a person's battery, and then recharge it, and recharge it again. But it is only when one has a generator of one's own that we can talk about motivation. One then needs no outside stimulation. One wants to do it. "*

According to Herzberg's theory, for a worker to be happy and productive, certain environmental factors must not cause discomfort. Although the elimination of these environmental discomforts may make a worker less dissatisfied, it will not necessarily motivate them. Herzberg believed that workers get motivated through feeling responsible for and connected to their work, not through rewards. In highly motivated individuals, the work

itself is rewarding. Managers who wish to apply this theory can give employees more authority over the job and offer direct individual feedback.

Herzberg argued that while underpaying an employee can serve as a dissatisfier, it does not mean that more money will increase job satisfaction or increased motivation (Kohn 1993). Kohn (1993) agrees: "Very few things threaten an organization as much as a hoard of incentive-driven individuals trying to curry favour with the incentive dispenser." Incentive systems can also mask underlying systemic organizational issues and stifle creativity since people are less inclined to take risks or explore possibilities because they are more concerned with what they will get for completing a task. Rewards are the enemy of exploration, according to Kohn (1993). Kohn (1993) says that over 70 studies show that rewards and disincentives are not only ineffective in the long-run, but are counterproductive. She believes this is because rewards distract from the natural tendency to learn, achieve, and satisfy natural curiousity (Maddock and Fulton 1998).

Katzell and Thompson (1990) have synthesized a vast body of research and theory regarding work motivation and have distilled the findings into seven "motivation imperatives:" a) Make sure that workers' motives and values are appropriate for the jobs on which they are placed, b) make jobs attractive to and consistent with workers' motives and values c) define work goals that are clear, challenging, attractive, and attainable, d) provide workers with personal and material resources that facilitate their effectiveness, e) create supportive social environments, f) reinforce performance, and g) harmonize all of these elements into a consistent focus on quality of work-life.

2.10.4 Summary of the motivational theories

Cognitive theories and non-cognitive theories are derived from very distinct views about human nature. Non-cognitive theories assume that behaviour can be influenced by manipulating rewards and incentives, while cognitive theories place more emphasis on the individual's ability to respond cognitively to stimuli. Yet, this research is not designed to debate the merits of either theory in influencing evaluation behaviours. In fact, some authors accept the notion that we can affect change if we apply managerial practices based on both cognitive and non-cognitive theories. They state that most organizational scholars now agree that several theories contribute to understanding of what influences the behaviour of an employee and should be applied in conjunction with each other (Sims and Lorenzi 1992).

3.0 STATEMENT OF THE PROBLEM

Goals for natural resource Extension in the United States have been expanding in recent years in response to changing public values (S. Reed, Director, Forestry Extension, Oregon State University, pers. comm., June 2002). In addition to their historic role in educating non-industrial private forestland owners, Extension natural resource professionals are also now focusing on broader public amenities such as watershed protection, wildlife, and other non-timber forest resources. More new natural resource Extension positions focus on water quality and other non-timber-related topics than ever before (M. Koelling, Professor of Forestry, Michigan Statue University, pers. comm., Sept. 2002). The relatively recent national Public Issues Education Initiative focuses on public education of community and natural resources (OSU Extension 2001).

The success of natural resource and agriculture Extension programs has historically been measured using indicators related to changes in knowledge, skills, attitudes, aspirations, and practices of private land owners for improvements related to the "benefit of themselves and society," with an emphasis on income-generation for the landowner (Hamilton and Biles 1998). However, as goals of natural resource Extension expand to reflect broader priorities, indicators of success have also expanded. Demonstrating success toward goals relating to public values requires more than accounting for the number of programs, dollars spent, numbers of participants, and practices and income associated with private land. Some measure of the protection and enhancement of public values is now an essential part of the evaluation equation. Developing indicators of success has been a challenge as these goals expand.

However, even before the expansion of goals that made evaluation more complex, the Cooperative State Research and Extension Education Service (CSREES) had already identified a gap between the need to document impacts and the practice of evaluation. An internal document from 1977 reported that evaluation at all levels of the organization had been "short on impacts and long on documenting participation and activity levels of programs" (Warner and Christenson 1984). In other words, reports included information on who participated and what they participated in, rather than the difference that their participation made vis-à-vis longer-term goals. The National Extension Committee on Organization and the Policy Task Force on Evaluation pointed to the tendency among managers to take their program's legitimacy for granted and only in crisis to "rally to save the organization" (Warner and Christenson 1984). The committees recommended that policies be developed to induce more program evaluation by Extension practitioners for local programs.

Some state-level Extension offices have begun to re-examine performance standards for their employees in light of increased demand for accountability (Cooper and Graham 2001). The number of core competencies identified for individual Extension practitioners, including competencies in program evaluation, has increased over time. Even as early as the

1970s, Extension evaluation was encouraged at all levels of the organization; local-level practitioners were encouraged to gain more knowledge and skills in program evaluation and demonstrate longer-term impacts through formal data collection, not just informal observation (Bennett 1976). Despite this early discussion, evaluation practices still remain highly variable across states.

The effectiveness of any organization depends on the motivational level of its employees (Lindner 1998). Since demonstration of long-term impacts of Extension programs is expected, then high levels of employee motivation to conduct program evaluation is paramount. According to Geering (1980), there are many interactive factors that explain the variation in performance level among employees. Intrinsic (personal) factors such as ability, personality, and background account for some differences. Extrinsic (external) factors such as organizational structure, the nature of the work, leadership style, and reward systems also affect employee performance levels. Across the U.S., there are several variations in the nature of natural resource Extension positions. For example, in some states, NREP positions are tenure-track while others are not, some NREPs have achieved tenure and some have not, and some NREP positions are partly or fully funded through grant ("soft") dollars while others are core-funded. These variations could be associated with different levels of motivation and behaviours for conducting evaluation.

No previous study has been conducted to describe NREP practices in program evaluation across the U.S. or to identify factors associated with their practice of program evaluation. Extension administrators have little information to guide them in creating rewards and disincentives, recruitment, performance standards, and training to enhance program evaluation conducted by NREPs. If managers create incentives and disincentives and other programs to boost evaluation practice in the absence of good information about what makes NREPs "tick" regarding program evaluation, they could be investing in the wrong places for the wrong reasons. Pardee (1990) suggests: "in order to understand how to motivate their subordinates, managers need to know what energizes human behaviour." A better understanding of what is behind NREPs program evaluation behaviours will help administrators place the "right" emphases in the "right" places to encourage and empower local NREPs to conduct higher level program evaluations.

4.0 RESEARCH HYPOTHESES

Motivational theories lead to a framework for examining the factors that influence program evaluation behaviour among NREPs. I hypothesized that the level of program evaluation conducted by NREPs is a function of intrinsic and extrinsic factors, and personal characteristics of NREPs. I surmised that positive attitudes toward program evaluation and perceived organizational support for evaluation, combined with various personal characteristics such as evaluation confidence levels, age, and years of experience are associated with levels of program evaluation conducted.

My eight research hypotheses are outlined in Table 4-1.

Table 4-1 Research Hypotheses

1) There is a positive linear relationship between "**attitude toward program evaluation**" and "level of evaluation conducted by NREPs most of the time in the past 12 months." The higher the attitude score of NREPs toward program evaluation, the higher the level of evaluation that has been conducted by NREPs most of the time in the past 12 months.
2) There is a positive linear relationship between "**perceived organizational commitment**" and "level of evaluation conducted by NREPs most of the time in the past 12 months." The higher the perceived commitment of the organization to evaluation, the higher the level of evaluation that has been conducted by NREPs most of the time in the past 12 months.
3) The level of evaluation conducted by NREPs most of the time in the past 12 months differs by **age** (older NREPs conduct higher levels of evaluation than younger NREPs).
4) The level of evaluation conducted by NREPs most of the time in the past 12 months differs by **years of experience** in the Extension profession. (NREPs with more years of experience conduct higher levels of evaluation than less experienced NREPs).
5) The level of evaluation conducted by NREPs most of the time in the past 12 months differs by **funding source**. (NREPs whose positions have a higher percentage of grant/soft funds conduct higher levels of evaluation).
6) The level of evaluation conducted by NREPs most of the time in the past 12 months differs by **position classification**. (NREPs whose positions are considered "tenure track" have conducted higher-level evaluations in the past 12 months than those whose positions are not "tenure track").
7) The level of evaluation conducted by NREPs in the past 12 months differs by **tenure status**. NREPs whose positions are "tenure track" but who have not yet achieved tenure status have conducted higher-level evaluations in the past 12 months than those whose positions are tenure-track but who have already achieved tenure status.
8) The level of evaluation conducted by NREPs in the past 12 months differs by **whether they believe their personnel performance appraisal is based on the basis of program evaluation behaviour**. (NREPs who believe their personnel performance is based on the level of program evaluation conducted conduct higher-level evaluations than those who don't believe their personnel performance is based on the program evaluation conducted).

5.0 RESEARCH METHODS

5.1 Research design

This study "described and interpreted conditions and relationships that exist" (Ary et al. 1996). It was impractical to conduct this study using traditional experimental design techniques because the variables occurred in a natural, not a manipulated environment. Independent variables were not controlled because NREPs were studied in their natural work environment.

I used a web-based survey using a commercial survey tool called "Zoomerang"[1] using followed guidelines outlined by Dillman and Salant (1994). The survey included questions relating to NREP characteristics as well as attitude and behaviour relating to program evaluation, including sources of motivation for conducting program evaluation. Questions relating to current and desired skill level in program evaluation and available resources were asked. Respondents were also asked in which U.S. state they work, but that variable was not used in the analysis because of potential breach of confidentiality (some states had only one or a few respondents) and because state of employment is not a factor related to the hypotheses.

Recent literature reveals both limitations and advantages of web-based surveys compared with mail, telephone or in-person surveys (Schonlau et al. 2002). While they are thought to be much faster than conventional survey modes, there is at best only marginal improvement in overall response times. Not all computer programs or computers can support the commercially available survey tools. Because the technology has the capability of limiting missing answers by forcing responses on certain key questions, I chose not to make most questions mandatory because of the risk that respondents would stop completing the survey. One typical shortcoming of web-based surveys is coverage error because it is expected that not all respondents have access to computers (Schonlau et al. 2002). In the case of this survey, however, all respondents had e-mail and web access through their work, but at least 12 individuals indicated difficulty in accessing or submitting the survey because of limitations in their systems. Five surveys were submitted from respondents by facsimile. Seven surveys were completed and submitted but were not recorded by Zoomerang. Some respondents who generally receive high numbers of e-mail messages may have deleted the survey before responding. Because "snail mail" is now less common in some cases than e-mail, a mail survey could be less likely to lead to non-response than a web-based survey. To counteract this, I sent reminders from my personal email address rather than the Zoomerang return address and addressed each recipient using their first names. Response rate increased significantly after the reminder was sent.

[1] Zoomerang is an on-line survey software created in 1999 by MarketTools, Inc. and can be accessed at www.zoomerang.com

5.2 Population and sample

The target population included 523 natural resource Extension practitioners in the 50 states of the U.S. who work under the Cooperative State Research and Extension Education Service (CSREES). The entire target population received a questionnaire and 224 county/region-based and campus-based completed and returned the questionnaire. Five respondents who reported having less than 20% of their work devoted to Extension were not included in the analysis, leaving 219 useable surveys at a response rate of 42%. Given the small size of the population, the low cost of sampling, and the ease of obtaining NREP email addresses, it was reasonable to conduct a census rather than a random sample of the population. The responses can be considered a random sample for the purposes of running the statistical tests; the risk of bias from non-responses in my study is the same as for a random sample, as indicated by a non-respondent survey described in 5.3.

In most states, NREPs included both campus-based and county-based personnel. I obtained an initial e-mail address list of those identified as working in forestry, urban forestry, or forest products utilization from the Cooperative State Research Extension Education Service (CSREES) website and confirmed the list by contacting each state's Extension administrative offices and for complete lists.

Given rapid changes in email and web use, it is difficult to determine a reasonable response rate for a web-based survey. While one would expect a very high response rate because of the ease and immediacy of the World Wide Web compared with telephone or mail surveys, it could be lower because of email overload, fears about viruses, and increased competition for respondents' attention due to increased email traffic. Changes in email and web culture and use occur rapidly, affecting the likelihood of response (such as "smarter" viruses that lead to greater caution by potential respondents in opening email messages from unfamiliar email addresses). For these reasons, there is not a widely accepted response rate for web-based surveys. However, a 42% response rate was considered very acceptable for this study.

5.3 Sampling non-respondents

To avoid non-response bias that can result when there is a difference between respondents and non-respondents, I randomly selected 30 non-respondents and conducted a brief email survey asking five demographic and job-related questions (see Appendix B) to aim for a 10% sample of non-respondents. I randomly selected the number 8 and surveyed every 10th name from the list of non-respondents (non-respondents numbers 8, 18, 28, etc). The first set of non-respondent surveys produced only seven responses. I followed up with a second personal email to 50 more randomly selected NREPs (every sixth name, starting with the randomly chosen number 2 (non-respondents numbers 2, 8, 14 etc.) that led to a total of thirty surveys returned. They answered questions about job classification, position classification, tenure status, percent of salary from soft funds, and years in Extension.

Thirty-four percent of respondents were tenure-track while 50% of non-respondents were tenure-track. Slightly more non-respondents (57%) than respondents (47%) were campus-based. Non-respondents received a lower portion of their salary from soft funds (mean of 15% compared with mean of 28%).

To determine the statistical difference between respondents and non-respondents for the categorical variables, I ran a Z-test to test differences between proportions for which a z-test is required (Fields 2000). With samples greater than 30, the t-test converges with the z-test. Since the z scores were each less than the critical z of 1.96, there were no significant differences between respondents and non-respondents for those variables. I also ran independent sample t-tests using the two continuous variables to determine if there was a significant difference in means between respondents and non-respondents with regards to percent of salary from grant ("soft") funds, and years in Extension (see Table 5-2 and Table 5-3). For years in Extension, I can assume equal variances ($p < .05$, Levene's test) and the significance was $p = .482$ which means there is no significant difference. However, for the percent of salary from grant ("soft") funds, I could not assume equal variances ($p > .05$, Levene's test), and the p was .043, which means that I could not reject the null hypothesis of no difference. In other words, respondents had a significantly higher percentage of their salary funded by soft funds than non-respondents. However, since significance was not high, and it was the only difference found between respondents and non-respondents, I proceeded on the assumption that statistical inferences could be made from the existing data.

Table 5-1 Differences between respondents and non-respondents–categorical variables

The top percent indicates respondent ratio; the bottom number represents the number of respondents selecting the option.	**Respondents**	**Non-respondents**	**Z scores**
Campus-based	47% 102	57% 17	-.83
Have tenure-track positions	34% 72	50% 15	-1.18
Have achieved tenure (of total)	28% 61	33% 10	-.24

Table 5-2 Summary statistics between respondents and non-respondents by years of Extension

	N	**Mean**	**Std. Deviation**	**t-value**	**Sig.**
Respondents	217	12.73	9.251	-.705	.482
non-respondents	30	14.03	10.473		

Mean difference = -1.29
Levene's test- equal variances assumed F= 1.451, p= .230

Table 5-3 Summary statistics between respondents and non-respondents by percent salary from grant ("soft") dollars

	N	Mean	Std. Deviation	t-value	Sig.
Respondents	217	28.24	40.070	1.696	.043
Non-respondents	30	15.33	30.398		

Mean difference = 12.90
Levene's test- equal variances not assumed *F= 11.618, p = .001*

5.4 Data collection methods

The Behavioural Research Ethics Board of the University of British Columbia issued a Certificate of Approval for my survey, cover letter, and research methodology in June 2003. I recruited evaluation specialists from the Extension Education Evaluation Topical Interest Group (EEE-TIG) of the American Evaluation Association and other U.S.-based Extension specialists to review the survey tool for face and content validity. Six individuals reviewed the draft and made comments for improving clarity. Their recommendations led to wording changes in the survey questions and instructions, such as the addition of questions pertaining to confidence and promotion and tenure. One reviewer recommended changing a question from "Is your position funded from soft funds or core funds?" to "What percentage of your salary is funded from soft (grant) funds?" to capture those whose salaries are partially funded by soft (grant) funds.

I obtained technical assistance in setting up the web-based survey and ensured that the database was compatible with the data analysis software (Statistical Package for Social Sciences – SPSS- Version 11.0). In October and November 2003, I conducted a pilot survey of 4-H Agents in Oregon using the Zoomerang survey tool, in cooperation with the state 4-H office. Fifty-four surveys were e-mailed to 4-H agents and 33 were completed and submitted after an introductory e-mail message from the 4-H Program Leader to each agent and two e-mail reminders from me. I made several wording changes (such as adding questions on "gut reaction," "greatest barriers" and "Which of the following most closely describes how you evaluate your programs most of the time?") to the national natural resources survey tool as a result of what I learned on the pilot study.

I obtained a mailing list from the CSREES 'Personnel in Forest Management and Wood Products' directory (http://www.reusda.gov/nre/forestry/fordirect.htm) and cross-checked the list and contact information with each state Extension Office or university websites between June and November 2003. Many states had significant additions. Because this was an electronic survey, I did not attempt to obtain postal mailing lists.

The survey consisted of 30 closed and open questions, including four multi-part questions. Two questions were 5-point Likert-scale index questions focused on perceived organizational commitment to evaluation and attitudes about program evaluation. Other questions related to barriers to program evaluation, behaviours regarding program evaluation,

sources of evaluation information, motivating factors, confidence levels, skills gaps, respondent characteristics, and recommendations for improving the evaluation practice within their organization. The survey questions are listed in Appendix A.

Closed questions provided data that lent themselves readily to quantitative analysis, while open questions provided depth of respondent viewpoints and experiences. Four open questions asked respondents to elaborate on a) evaluations they have conducted other than end-of-event questionnaires, b) barriers to evaluation, c) recommendations to improve the practice of evaluation in their organizations, and d) other comments. All responses are included in appendices C-E. Four questions that represented key variables were "mandatory;" in other words, respondents could not submit the questionnaire until those questions were answered. The mandatory questions were selected to ensure a balance between ensuring critical data and risking that respondents would abandon the questionnaire. Mandatory questions included a) whether or not they conducted evaluation in the past twelve months, in what state they work, their job classification, and highest education level. (The variable, "how they evaluate programs most of the time," should have been mandatory, but this was inadvertently left out. Fortunately, only six respondents chose not to answer this key question).

I used a summated rating scale (Spector 1992; Ary et al. 1996) to measure two of the independent variables (attitude and perceived organizational commitment). Summated rating scales are multivariate measures used to assess complex psychometric properties such as perceptions and attitudes. When responses to multiple sub-questions (which can also be called indicators) about a single concept such as perceived organizational commitment to program evaluation are averaged, they provide a more reliable indicator of the concept (Spector 1992; Ary et al. 1996). Survey question 7 (perceived organizational commitment) had 6 items in the scale and survey question 14 had 11 items. The summated scores for survey question numbers 7 and 14 were obtained by summing the scores of the sub-questions then averaging the values of the responses.

Cronbach's alpha coefficient was used to quantify the magnitude of intercorrelation or internal consistency of summated rating scales (where 1.0 means a perfect correlation among items in the scale. The closer to 1.0, the better the scale indicates the concept being tested). The alpha coefficient was .58 for survey question 14 and .77 for survey question 7. By removing three items in the index of question 14 (sub-questions 14g, 14h, and 14j), the reliability alpha increased to .78. Thus, I recoded summated survey question 14 and used the new variable in my analyses using the remaining 7 items in the index.

For questions 7 and 14, I mixed unfavourable statements and favourable statements to ensure the most accurate results (requiring respondents to read and consider each question carefully). I reversed the unfavourable scores during compilation since "disagreement" in a Likert-type scale is assumed to be equivalent to "agreement" with favourable statements (Spector 1992; Ary et al. 1996). For survey question 7, I reversed two of six sub-questions,

and in survey question 14 I reversed six of 10 sub-questions such that a score of "1" was reversed with "5" and "2" was reversed with "4." A score of "3" in both questions was left unchanged since it was a neutral score.

Additionally, I contacted each state's land-grant university Human Resources Office and/or Extension administrative office to obtain a blank copy of the personnel appraisal forms that included criteria for personnel appraisal. Some forms were available on a website but most were e-mailed or mailed to me. I was able to obtain information on performance appraisal criteria from 39 states. I reviewed each of 39 state's criteria to determine to what degree conducting program evaluation was present among the criteria.

5.5 Limitations, assumptions, and sources of potential error

There are four types of errors common in surveys: coverage error, sampling error, measurement error, and non-response error (Dillman and Salant 1994). To avoid the errors, I took the following precautions:

- **Coverage error:** I used the most up-to-date e-mail list available from the national Cooperative State Research, Education, and Extension Service database, and field checked the list by contacting each state for an updated list. I checked the mailing list for duplicate entries. I surveyed non-respondents to look for a pattern that could lead to bias in my analysis.

- **Sampling Error:** Sampling error refers to the error that occurs when less than 100% of the population is surveyed (unless it is a census where all members of the population are surveyed). To minimize sampling error, I surveyed all possible respondents in the population and sent them repeated reminder notices. The final reminder notice was addressed to them by their first name (rather than a generic "Dear colleagues"); the response rate increased from about 10% to over 40% after the last reminder.

- **Measurement error:** Measurement error can come from two sources: the survey method used and the questions themselves (Dillman and Salant 1994). To avoid this type of error, a) I used both open-ended questions and closed-ended questions to provide insight regarding why respondents selected a particular answer; b) I had a panel of experts review the questions for face and content validity; c) I pre-tested the survey with 33 respondents who were similar to the target population for the national survey, d) I used simple, clear, short sentences that avoided jargon, and e) included definitions of unfamiliar terms.

- **Non-response error:** The size of non-response error is a function of two factors: response rate and extent to which respondents differ from non-respondents (Wiseman 2003). Non-response error is a potential problem if response rate is low or if respondents and non-respondents differ. I compared non-respondents to respondents to see if there was a difference among their characteristics. I aimed for the highest possible response rate through two reminder notices with the survey attached to non-respondents and by providing an incentive to respond (a draw for smoked salmon to 10 randomly selected respondents).

In addition, social desirability error was possible on questions relating to attitude about program evaluation. One respondent voiced concern about anonymity; this fear could also lead to social desirability error if respondents answered questions based on how they felt they should feel or behave, not how they actually feel or behave. There is also risk of sample bias because those who are most motivated to conduct evaluation have been most likely to complete the survey.

As a national study with highly variant numbers of NREPs per state, it was not meaningful to report on the results by state. Doing so also could also have risked loss of anonymity in cases where states had only a few NREPs. Some states had high numbers of NREPs (such as Colorado at 75 NREPs) and some states, such as South Dakota, had only one NREP.

There was also some variability in how each state defined "natural resource Extension" and the Extension function is organized. In most cases, I had to rely on state definitions because they provided me their lists. As a result, it was difficult to identify the population precisely and it was highly heterogeneous.

The survey was sent to all NREPs rather than a sample of NREPs, so the responses do not represent a random selection of cases. This could provide a source of bias in the results because random sample is an assumption of analysis of variance and multiple linear regression statistical tests. However, the randomly selected non-respondent survey respondents were similar to the respondents in all but one factor so there is little evidence of non-response bias.

5.6 Quantitative data analysis

I used SPSS Version 11.0 software to code and analyse the data. I used four procedures:

- Descriptive statistics
- ANOVA
- Cross tabulations with chi-square correlation statistics
- Multiple linear regression

In addition, I conducted a cluster analysis to classify the responses into similar groups, as described in Section 6.7.1. The descriptive statistics including means, medians, modes, ranges, standard deviations, and standard errors of mean of continuous data such as "age"

and "number of years in Extension" and frequencies and percents of categorical data such as "confidence level" and "tenure status" of respondents. Two survey questions included sub questions whose values were averaged and analysed as interval summated data: "attitude," (survey question 14) "perceived organization commitment," (survey question 7).

Parametric techniques (ANOVA, multiple linear regression, and independent sample t-tests) and non-parametric tests (chi-square) were used to test for differences and correlations among the variables. A linear regression was attempted and five variables had linear relationships with evaluation behaviour but only a small portion of the variance in evaluation behaviour could be explained by the variables. A cluster analysis was conducted to group respondents into three clusters based on the level of program evaluation conducted (survey question 15). Table 5-4 summarizes the statistical tests and procedures used.

Assumptions and diagnostic tests of assumptions for ANOVA and multiple linear regression are discussed in Section 6.7.3 (multiple linear regression) and Section 6.7.4 (ANOVA).

In this study, two variables were used measure evaluation behaviour: survey question 15 (frequency of program evaluation conducted) that is considered a categorical variable for the analysis, and survey question 16 (level of program evaluation conducted most of the time) that is considered a continuous variable for the analysis. In the regression, "frequency of program evaluation conducted" is used as the dependent variable. In the ANOVA, however, "level of program evaluation conducted most of the time" is used as the dependent, or grouping variable.

While the variable, "level of program evaluation conducted most of the time" (survey question 16), could be considered ordinal because it represents a ordered ranking of program evaluation behaviour, it was also used as a categorical variable for the analysis.

Table 5-4 Variables and statistical tests used

Attitude toward program evaluation (Summated continuous scale - survey question 14)	Level of evaluation conducted by NREPs most of the time (categorical scale - survey question 16 as grouping variable)	ANOVA
Perceived organizational commitment (Summative continuous scale - survey question 7)	Level of evaluation conducted by NREPs most of the time (categorical scale - survey question 16 as grouping variable)	ANOVA
Age (continuous scale - survey question 25)	Level of evaluation conducted by NREPs most of the time (categorical scale - survey question 16 as grouping variable)	ANOVA

Years in Extension (continuous scale - survey question 23)	Level of evaluation conducted by NREPs most of the time (categorical scale - survey question 16 as grouping variable)	ANOVA
Percentage of salary derived from grant funds (categories) (recoded into categorical scale because distribution not normal- survey question 28)	Level of evaluation conducted by NREPs most of the time (categorical scale - survey question 16)	Cross tabulation with chi-square test for association
Position classification (tenure/non-tenure track) (categorical scale - survey question 26)	Level of evaluation conducted by NREPs most of the time (categorical scale - survey question 16)	Cross tabulation with chi-square test for association
Tenure status (of tenure-track respondents) (categorical scale - survey question 27)	Level of evaluation conducted by NREPs most of the time (categorical scale - survey question 16)	Cross tabulation with chi-square test for association
Performance appraisal criteria including whether or not conduct program evaluation (categorical scale - survey question 9	Level of evaluation conducted by NREPs most of the time (categorical scale - survey question 16)	Cross tabulation with chi-square test for association
Age, years of experience, position classification, tenure status, perceived organizational commitment, attitude about evaluation, believes performance appraisal is based on program evaluation behaviour (continuous and dummy)	Frequency of program evaluation conducted (continuous scale - summated survey question 15)	Multiple linear regression
Frequency of program evaluation conducted (continuous scale - summated survey question 15)	Access to evaluation specialists (categorical scale - grouping variable- survey question 10g)	t-test
Frequency of program evaluation conducted (continuous scale - summated survey question 15)	Believes performance is assessed on basis of program evaluation behaviour (categorical scale - grouping variable- survey question 9)	t-test
Frequency of program evaluation conducted	Tenure status	t-test

(continuous scale - summated survey question 15)	(categorical scale - grouping variable- survey question 27)	

Table 5-5 shows the timeline of activities for the study. The survey was drafted in May 2003, received Human Subjects Committee approval in June 2003, was reviewed by the panel of experts in August and September, and launched in November 2003. I conducted a non-respondent survey in March-April 2004, and analysed and wrote the results in April-June, 2001.

Table 5-5 Timelines

Obtained copies of personnel appraisal forms from 39 states	2002-2003
Comprehensive examination	February 2003
Received ethics review approval from Human Subjects Committee	June 2003
Obtained database for population	May-November 2003
Recruited experts from the Extension Education Evaluation Topical Interest Group (TIG) of the American Evaluation Association to review the survey draft. Suggestions incorporated into survey	August-September 2003
Obtained technical assistance in setting up web-based survey	October 2003
Conducted pilot survey of Oregon 4-H agents and submitted report to state 4-H office	October-November 2003
E-mailed survey with cover letter to potential respondents	November 30, 2003
E-mailed first reminder notice to non-respondents with personalized cover letter	December 8, 2003
Emailed second reminder notice to non-respondents with hot link to survey	December 15, 2003
Web-based survey closed	January 9, 2004
Random draw for prizes (smoked salmon). Prizes mailed to winners	January 10, 2004
Conducted non-respondent survey	March-April 2004
Analysed results using SPSS and drafted dissertation	January 10-May 2004
Submitted draft dissertation to committee	June 2004

6.0 QUANTITATIVE RESULTS

This study was designed to test hypotheses related to program evaluation behaviour of natural resource extension professionals in the United States. A survey was conducted to assess factors related to attitudes and behaviours of program evaluation such as perceived organizational commitment to evaluation, sources of motivation, and attitudes towards evaluation. The survey was conducted using a web-based format that was made available to potential respondents by email.

6.1 Survey response

Invitations were emailed to 570 potential respondents, and 47 were returned because they were outdated email addresses. Response rate per state ranged from zero percent to 100 percent.

Seventeen individuals who did not complete the survey replied with reasons that they did not intend to complete the survey. Of those, fifteen respondents said that the survey was not relevant to them, and two said that they did not have time to complete it. Eleven surveys were eliminated from the dataset because the respondents had less than .20 FTE in Extension. 276 non-respondents did not provide any reason for not responding. The analysis was conducted using 219 usable surveys.

Table 6-1 Survey response rate

Population size (N)	523
Number of useable returned surveys (n)	219
Number of non-respondents	304
Response rate (percent)	**41.8**

Table 6-2 Reasons for non-response/removal from dataset

Not relevant to them	15
No time to complete	2
Less than .20 FTE in Extension	11
No reason given	276
Total non-respondents	**304**

6.2 Respondent characteristics

6.2.1 State in which respondent was employed

More questionnaires were completed from respondents in the state of Colorado than any other state. There were 36 submitted from Colorado, 19 from Michigan, 13 from Pennsylvania, 11 from New York, and 10 from Oregon. Other states had fewer. The table in Appendix G show the response data by state (number of potential and actual respondents, percent response rate per state).

6.2.2 Current employer

Forty-seven percent (102 respondents of 219) considered themselves employed by "Land Grant University-county or regional-based" and 50 percent (110 respondents) considered themselves employed by "Land Grant University-campus based." Four respondents (slightly more than one percent) who selected government agencies were from Colorado (the other 32 Colorado respondents selected Land Grant University-county-based as their employer, reflecting the partnership between Colorado state government and the Land Grant University). Three individuals selected "other" and one of those specified "adjunct faculty at Land Grant University."

Table 6-3 Current employer

	Frequency	Percent
Land Grant University- campus based	102	47
Land Grant University- county/region based	110	50
Government agency	4	2
Other	3	1
Total	**219**	**100**

6.2.3 Age of respondents

Of the 210 respondents who provided their age, the average was 45 years with a standard deviation of 9.6 years. The oldest respondent was 72 years old and the youngest was 25 years old. The most common ages were 47 (13 respondents), 45 (10 respondents), and 39 (10 respondents). Nine individuals chose not to disclose their age. Age was normally distributed.

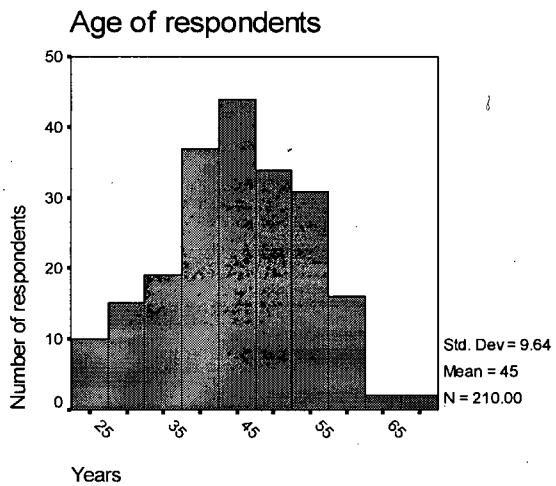

Figure 6-1 Age of respondents

OptionButton1

6.2.4 Years in extension

The average number of years of service in Extension ranged widely among the 217 respondents. The average was 12.7 years with a standard deviation of 9.3 years. The median was 10. Two persons were new to Extension, while the respondent with the longest service had served for 37 years. The most common number of years was five. The distribution was positively skewed.

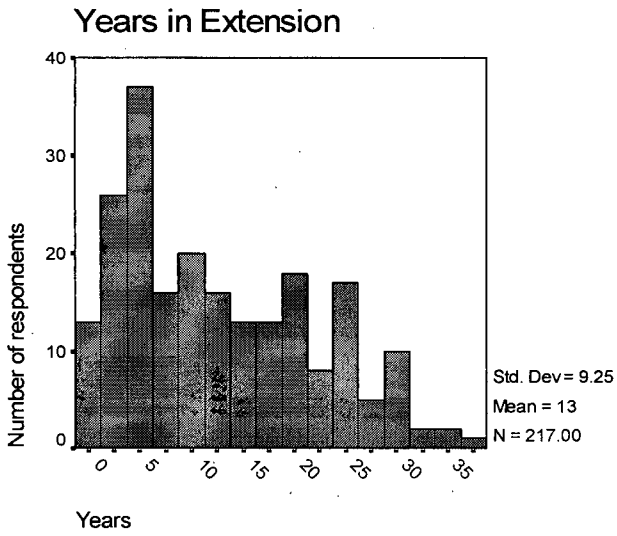

Figure 6-2 Years in Extension

6.2.5 Position classification (tenure track/not tenure-track/professional staff)

Thirty-four percent of 216 respondents (72 respondents) had tenure-track positions and 24 percent (52 respondents) was non-tenure track faculty. Forty percent (86 respondents) was non-faculty professional staff.

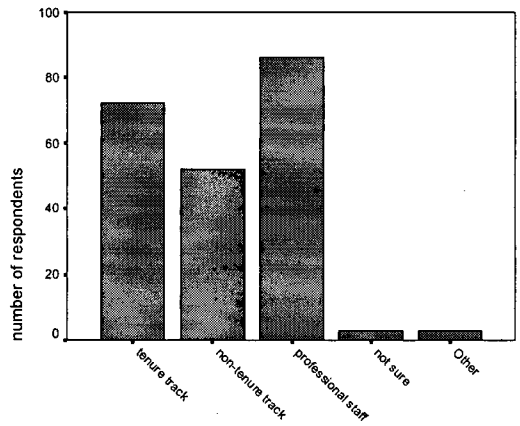

Figure 6-3 Position classification

6.2.6 Job classification

Of 219 respondents, thirty percent (74 respondents) were county or region based extension agents, and 39% percent (96 respondents) were specialists with state-wide responsibilities. Two percent (five respondents) were program assistants and 15% (37 respondents) were state forestry employees, and three percent (7 respondents) were other.

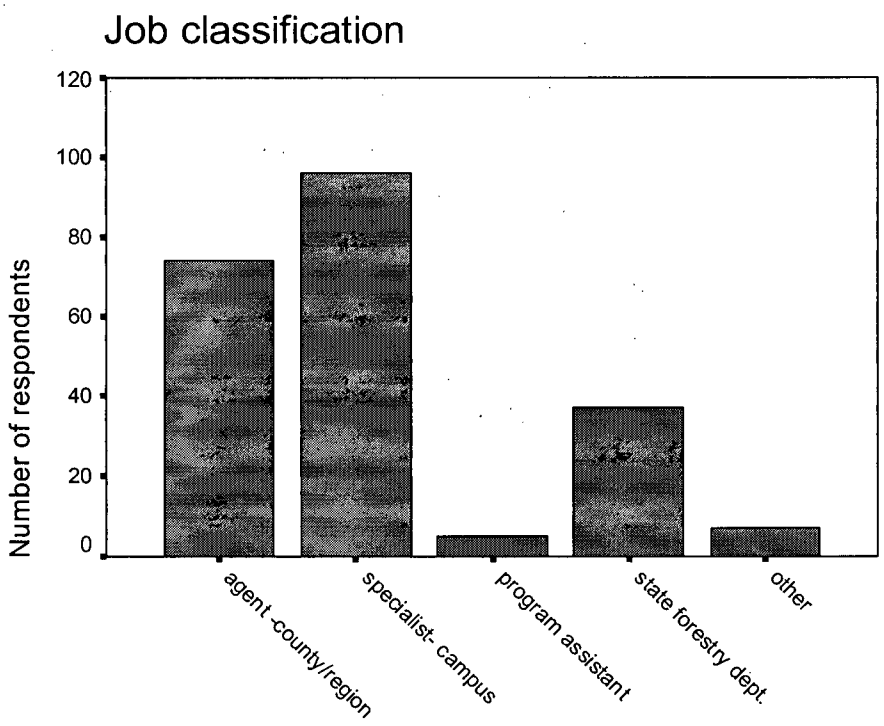

Figure 6-4 Job classification

6.2.7 Tenure status of tenure-track faculty

Twenty-eight percent (61 respondents) said that they had achieved tenure and 11 percent (24 respondents) said that they had not yet achieved tenure status. One person was unsure. Figure 6-5 shows the percent of respondents and their tenure status. About three times as many tenure-track faculty respondents had achieved tenure than those who had not (61 compared with 24).

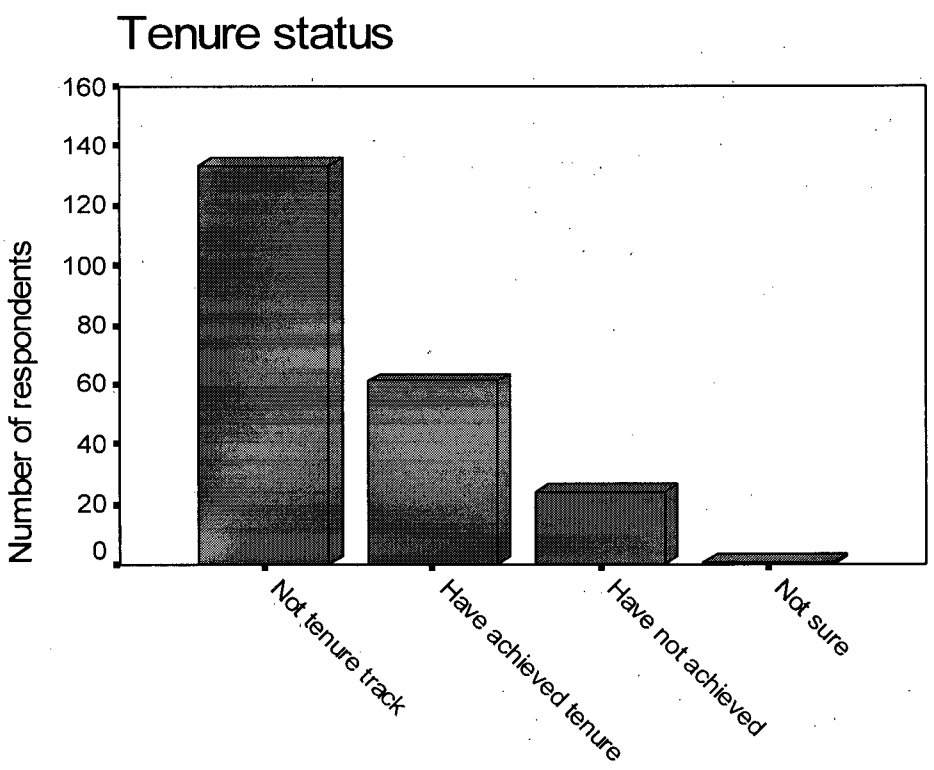

Figure 6-5 Tenure status

6.2.8 Highest education level

Forty-nine percent of respondents (108 individuals) said their highest education level was a Master's degree. Nineteen percent (42 respondents) said that a Bachelor's degree was their highest education level, while 31 percent (68 respondents) said that a PhD degree was their highest education level (Figure 6-6). One respondent selected high school degree as their highest education level and no respondents selected technical degree.

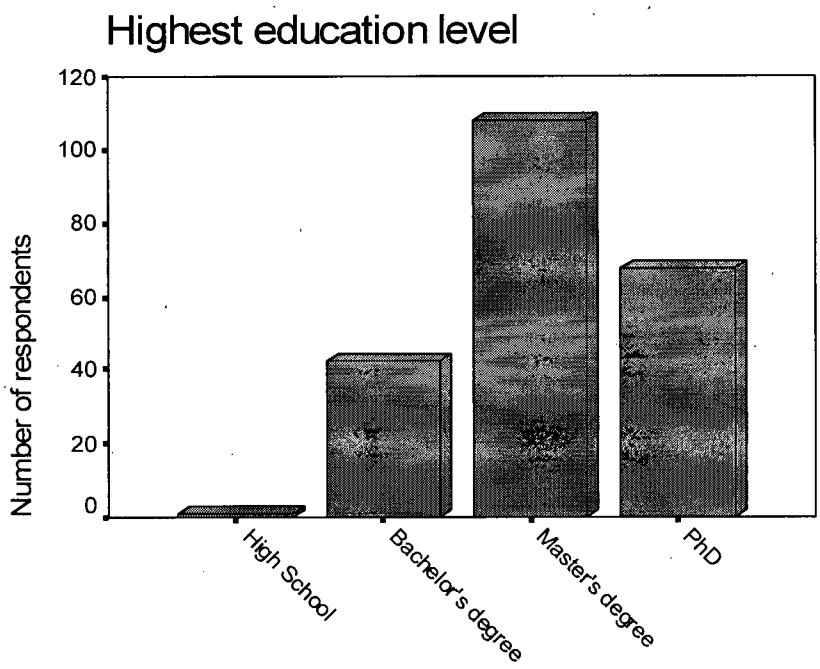

Figure 6-6 Highest education level

6.2.9 Source of funds for salary

Fifty-four percent (115 respondents of 217) said that none of their salary came from grant dollars, implying that a slight majority of positions is 100 percent core-funded. Seventy-six percent of respondents said that fifty percent or less of their salary comes from grant dollars. Nineteen percent (43 respondents) said that 99 or 100 percent of their salary comes from grant dollars. Two respondents didn't answer the question.

The mean percent of salary that comes from grant dollars is 28 but the median is 0[1]. The distribution is bi-modal. The standard deviation is 40 percent. The maximum was 100 and the minimum was 0. Given the non-normal distribution of this data, the median and mode are more meaningful than the mean.

The responses to this question show that the majority of respondents' positions were core funded and that almost half had some part of their salary derived from soft funds. The bimodal distribution reflects the current nature of Extension positions where most are either all core-funded or all-grant funded.

[1] More than half of respondents selected 0

Table 6-4 Percent of salary from grant ("soft") dollars (frequency)

Percent of salary	Number of respondents	Percent of respondents	Cumulative Percent
0	115	54	53
4	1	.5	54
5	5	2	56
6	1	.5	56
8	1	.5	57
10	10	5	61
15	4	2	63
20	6	3	66
25	7	3	70
30	4	2	71
40	2	1	72
50	9	4	76
60	2	1	77
75	2	1	78
80	3	1	79
90	1	.5	80
95	1	.5	80
99	16	7	88
100	27	12	100
Subtotal	217	100	
Missing	2		
Total	**219**		

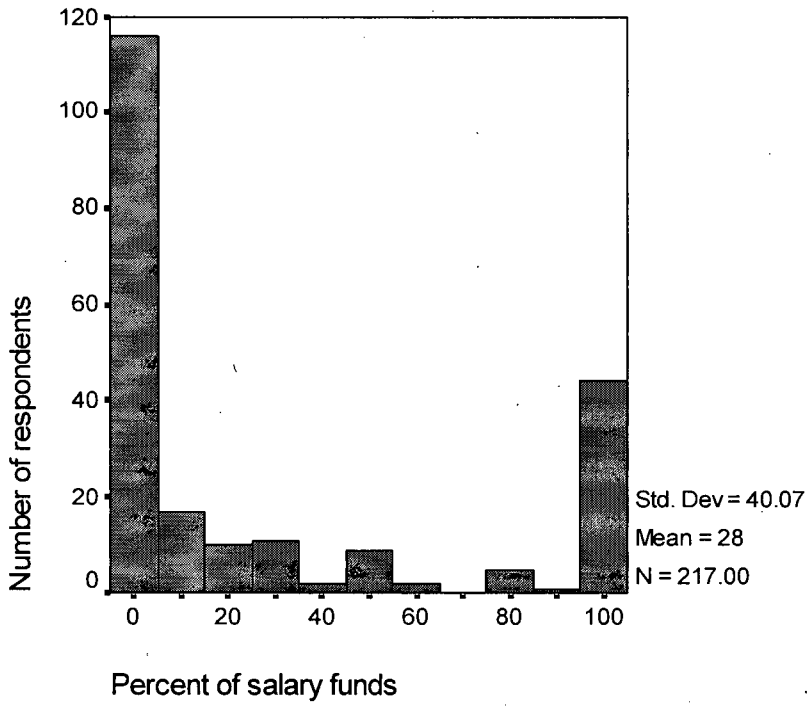

Figure 6-7 Percent of salary funds from grant (soft) dollars

Table 6-5 shows the percent of salary funded by grant dollars for each category of Extension position. The percent of agent and state-wide positions that are 100 percent grant-funded are not substantially different (65 and 58 percent, respectively), but a much smaller percentage of state forestry department employees are 100 percent core-funded (30 percent). A higher number of state forestry department employees are 100 percent grant-funded than the other two categories.

Table 6-5 Funds for salary from grant dollars * Job classification Cross tabulation

% of salary	county/regional agent	state-wide specialist	program assistant	state forestry dept.	other	Totals
0 %	47 64.4%	55 57.9%	1 20.0%	11 29.7%	1 14.3%	115 53.0%
1-50 %	11 15.1%	22 23.2%	2 40.0%	14 37.8%	1 14.3%	50 23.0%
51-99 %	7 9.6%	10 10.5%		6 16.2%	2 28.6%	25 11.5%
100 %	8 11.0%	8 8.4%	2 40.0%	6 16.2%	3 42.9%	27 12.4%
Total	73 100.0%	95 100.0%	5 100.0%	37 100.0%	7 100.0%	217 100.0%

6.2.10 Percent of job in Extension, service forestry or landowner assistance

This survey question was used as a filtering question to establish a sampling threshold (respondents with less than 20 percent in Extension, Service Forestry, or Landowner Assistance were removed from the dataset for analysis). After respondents who had less than 20 percent in Extension were removed, over half of the remaining respondents (54 percent or 119 respondents) had positions of 100 percent in Extension, Service Forestry, or Landowner Assistance. Eight-eight percent had 50 percent or higher. The mean was 85 percent and standard deviation was 21.5 percent. The range was 85 percent with 20 percent as the minimum and 100 percent as the maximum. The distribution was negatively skewed. The median was 100 percent but the mean was 85 percent.

Table 6-6 Percent of job in Extension (frequency)

Percent of job	Frequency	Percent
20	4	2
25	4	2
30	4	2
40	1	.5
50	13	6
51	1	.5
60	8	4
64	1	.5
65	3	1
67	1	.5
69	1	.5
70	7	3
75	17	8
79	1	.5
80	16	7
81	1	.5
85	4	2
90	7	3
95	5	2
100	119	55
Subtotal	218	100
Missing	1	
Total	**219**	

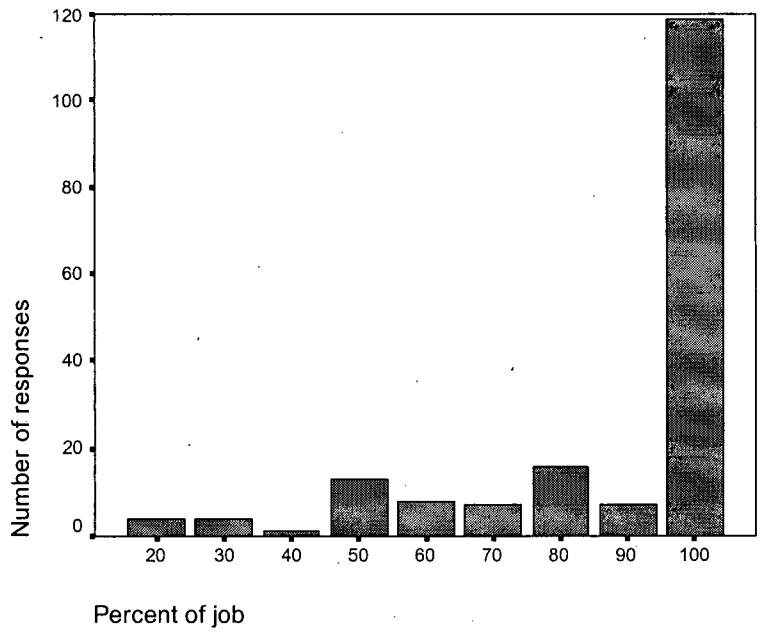

Figure 6-8 Percent of job in Extension

6.3 Perceptions towards program evaluation

6.3.1 "Gut" reaction toward evaluation

Sixty percent of respondents (130 individuals) said that they don't mind conducting evaluation and 27 percent said that they prefer to ignore it. Only two percent said that they absolutely dread it, and two percent doesn't think about it. Nine percent of respondents said that they "love doing it."

Table 6-7 "Gut" reaction towards evaluation

	Frequency	Percent
I love doing it	19	9
I don't mind it	130	60
I would prefer to ignore it	59	27
I absolutely dread it	5	2
I don't think about it	5	2
Missing	1	-
Total	**219**	**100**

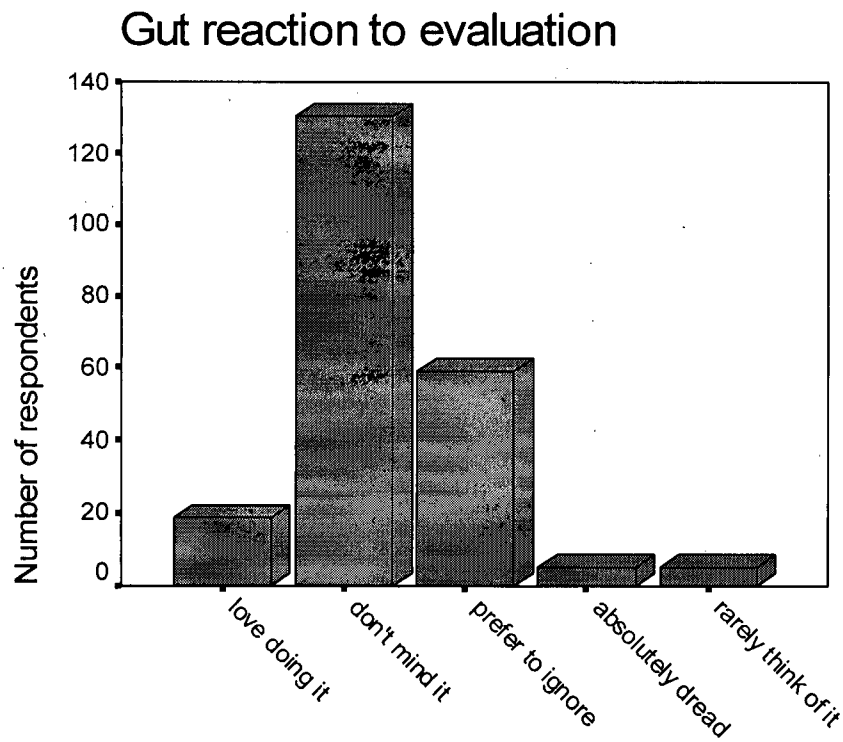

Figure 6-9 Gut reaction to evaluation

6.3.2 Attitude about program evaluation

Respondents were asked their opinion about program evaluation using 10 five-point index questions. The majority (sixty-five percent or 140 respondents) agreed or strongly agreed that program evaluation is an important part of Extension. Eighty-three percent (180

respondents) agreed that using evaluation results improves Extension programs. Responses varied regarding the statement that 'evaluation takes away from the "real" work of Extension;' Forty-nine percent of respondents strongly agreed, agreed, or were neutral regarding the statement that evaluation takes time away from the "real" work of Extension, while 51 percent disagreed or strongly disagreed.

Sixty-five percent of respondents believed that personnel appraisals should include criteria related to program evaluation, but a quarter was uncertain. Nine percent of respondents believed that performance appraisals should not include criteria related to program evaluation. Nearly a quarter of respondents agreed or strongly agreed that there is no consequence for not conducting program evaluations while a quarter was uncertain. Fifty-eight percent of respondents said that that they would conduct program evaluations even if it was not required. Nineteen percent would not. Thirty-seven percent of respondents believed that Extension professionals are not recognized for conducting program evaluations and 36 percent were neutral. Sixty-four percent disagreed or strongly disagreed with the statement that they rely on others in the organization to worry about program evaluation.

Thirty-eight percent (82 respondents) agreed or strongly agreed that they usually drag their feet about doing program evaluations.

On average, respondents felt the strongest about the statement that performance appraisals should not include criteria related to program evaluation; responses averaged 3.7 on a scale of 1-5, where five is strongly disagree. The average response to the statement "I usually rely on others to conduct evaluation" was also strongly held; respondents generally disagreed (average 3.7). The greatest diversity in responses occurred with the statements "evaluation takes time away from 'real' Extension work," "I usually drag my feet about doing program evaluations," "would conduct evaluation even if not required," and "there's no consequence of NOT doing program evaluation."

Table 6-8 Percent and frequency of attitudes about program evaluation

The top percent indicates total respondent ratio; the bottom number represents the number of respondents selecting the option	**1 Strongly Agree**	**2 Agree**	**3 Neither Agree nor Disagree**	**4 Disagree**	**5 Strongly Disagree**	**Percent**
1. Believes that program evaluation is important part of work	15% 32	50% 108	21% 45	14% 30	1% 2	100% 217
2. Usually drags feet about doing evaluations	6% 12	32% 70	31% 67	27% 59	4% 8	100% 216
3. Believes results of program evaluation improve programs	29% 62	54% 118	11% 23	5% 11	1% 3	100% 216
4. Doing program evaluation takes time away from "real" work of Extension	3% 7	22% 48	24% 51	42% 92	9% 19	100% 217
5. Performance appraisals should not include criteria re: program evaluations	1% 2	8% 17	26% 56	57% 124	8% 18	100% 217
6. Believes conducting program evaluation contributes to meaning of work	16% 34	57% 122	21% 46	5% 11	1% 2	100% 215
7. Believes there's no consequence of NOT doing evaluation	2% 5	20% 43	26% 56	44% 96	8% 17	100% 217
8. Extension professionals not recognized for conducting evaluations	6% 14	31% 67	36% 78	25% 55	1% 3	100% 217
9. Would conduct program evaluation even if wasn't expected	13% 27	45% 97	23% 49	17% 36	2% 5	100% 214
10. Relies on others to worry about program evaluation	0% 1	14% 29	22% 47	48% 102	16% 35	100% 214

Table 6-9 Attitudes about program evaluation

(five-point scale: 1 = strongly agree and 5 = strongly disagree)

	Considers program evaluation important part of Extension work	Usually drags feet about evaluation	Believes that using of program evaluation results improves programs	Evaluation takes time away from the "real" work	Performance appraisals should not include criteria related to evaluations
Mean	2.4	2.9	2.0	3.3	3.7
Median	2	3	2	4	4
Mode	2	2	2	4	4
Std. Dev.	.93	.99	.85	1.0	.78
Std. Error	.063	.067	.057	.068	.053
Range	4	4	4	4	4
Minimum	1	1	1	1	1
Maximum	5	5	5	5	5

(con't)

	Believes conducting evaluation contributes to meaning of work	Believes there's no consequence of NOT doing evaluation	Extension professionals in state not recognized for conducting evaluation	Would conduct program evaluation even if not expected	Relies on others to worry about program evaluation
Mean	2.2	3.4	2.9	2.5	3.7
Median	2	4	3	2	4
Mode	2	4	3	2	4
Std. Dev.	.79	.97	.93	.99	.92
Std. Error	.054	.065	.063	.063	.067
Range	4	4	4	4	4
Minimum	1	1	1	1	1
Maximum	5	5	5	5	5

The ten survey sub-questions of survey question 14 were summated into an index scale by adding the scores and averaging them. However, first Cronbach's alpha was used to measure internal reliability of the items in the index. Cronbach's alpha for all 10 sub questions was .58 (where 1.0 indicates that the items in the index are measuring the same thing). By eliminating three of 10 sub questions, Cronbach's reliability measure increased to .77. Following Spector (1992), attitude summated score was then calculated on the basis of the seven remaining sub questions. The following table is the attitude score after 3 items were removed to improve reliability of the scale (the remaining seven sub questions of survey question 14 are listed in Appendix A).

The mean attitude score was 2.47, and the mean was 2.42. Standard deviation was .59. The maximum was 4.43 and the minimum was 1.29.

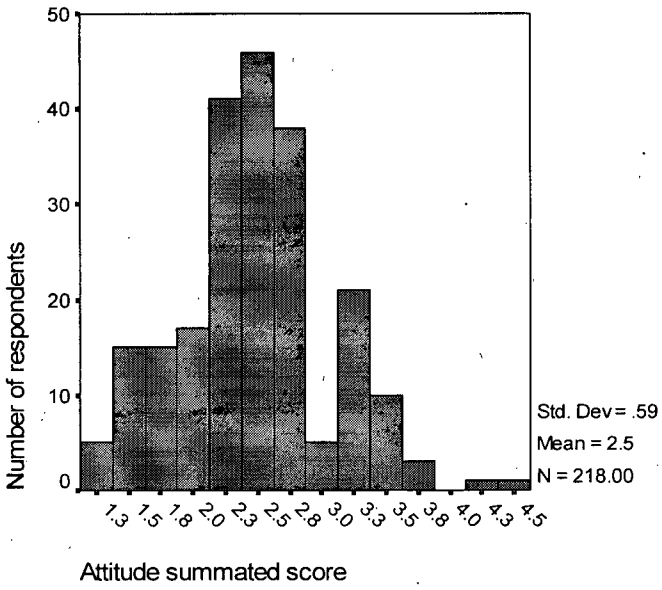

Figure 6-10 Attitude towards program evaluation

6.3.3 Motivation based on external factors

Respondents were asked to estimate what percent of their motivation to conduct program evaluation resulted from external and internal factors, with the total adding up to 100 percent. The highest number of respondents gave a 50-50 answer, but there was wide variation among the responses. On average, respondents are slightly more motivated by external factors such as promotion and tenure, grant reporting requirements, and performance appraisals than by internal factors such as desire to know if teaching is effective or desire for professional improvement, but this varied widely among respondents. Twenty-six percent of respondents (57 individuals) said the mix was 50-50. Four percent (nine individuals) said that the motivation was 100 percent external, and two percent (four individuals) said that the motivation was 100 percent internal. Nine percent (20 individuals) said that 75 percent of their motivation was external. The mean was 54 percent, the median and mode were 50 percent, and the standard deviation was 26 percent. The maximum was 100 percent and the minimum was 0.

Table 6-10 Motivation for evaluation due to external factors

Percent of motivation	Frequency	Percent of respondents	Cumulative Percent
0	4	20	2
2	1	.50	2
5	3	1	4
10	8	4	8
11	1	.50	8
15	3	1	9
20	13	6	15
25	8	4	19
30	11	5	24
35	2	.9	25
40	9	4	29
50	57	26.3	55
60	17	8	63
65	1	.50	63
70	15	7	70
75	20	9.2	79
80	16	7	87
85	3	1	88
90	10	5	93
95	3	1	94
99	4	2	96
100	9	4.1	100
Subtotal	217	100	
Missing	1		
Total	218		

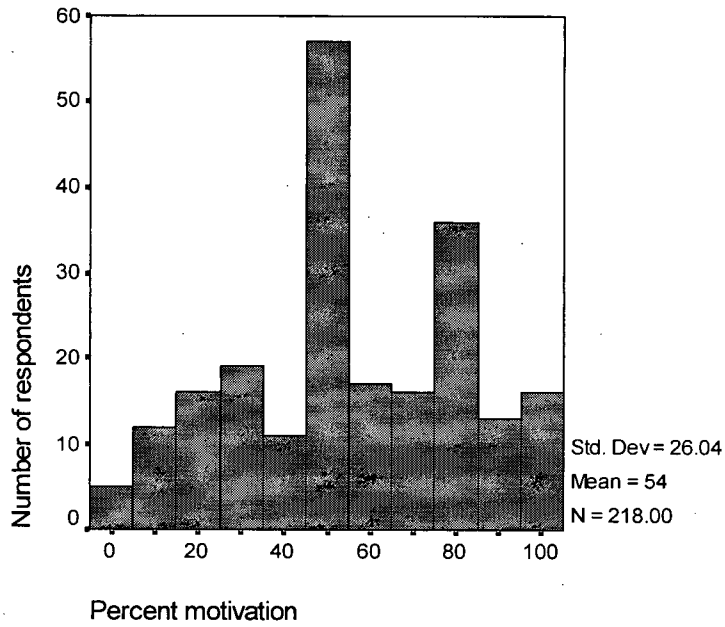

Figure 6-11 Percent motivation due to external factors

6.3.4 Confidence level

There was substantial variation in confidence levels for conducting program evaluations among respondents. Twelve percent (25 respondents) felt they were very confident and 42 percent (92 respondents) said they were moderately confident. Sixteen percent (35 respondents) said that they were moderately unconfident, and 28 percent (60 respondents) were neutral. Three percent (five respondents) said they were very unconfident. Two respondents did not answer the question.

Table 6-11 Confidence level in conducting program evaluations

	Frequency	Percent
Very confident	25	12
Moderately confident	92	42
Neither confident nor unconfident	60	28
Moderately unconfident	35	16
Very unconfident	5	2
Subtotal	217	
Missing	2	
Total	**219**	**100**

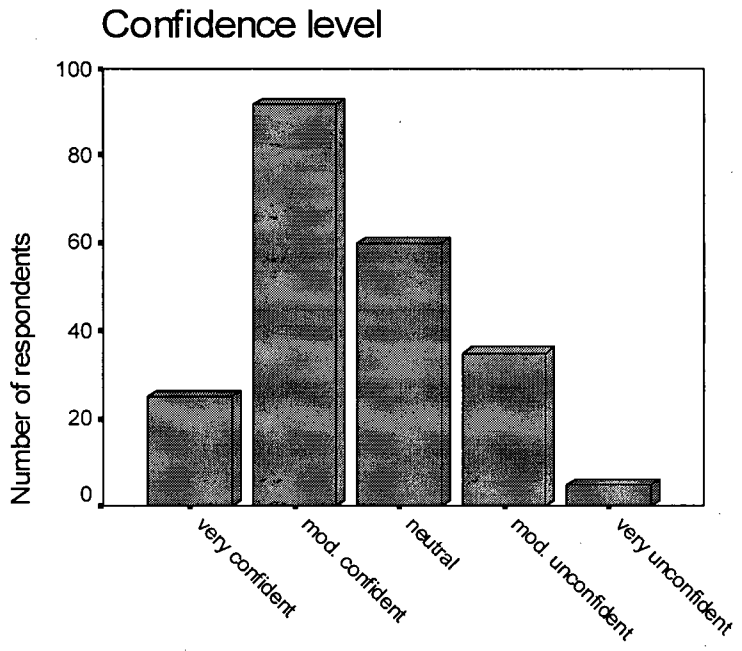

Figure 6-12 Confidence level in conducting program evaluations

6.3.5 Performance appraisal and evaluation

Responses were nearly evenly split between those who believed that their performance was assessed in part on whether or not they evaluated their programs, and those who believed it

was not. Forty-two percent said they believed it was, and 44 percent said that it was not. Fourteen percent (31 respondents) were not sure.

Table 6-12 Perceptions regarding performance appraisal

	Number	Percent
Believe performance is assessed on whether or not they conduct program evaluation	92	42
Believe performance is NOT assessed on whether or not they conduct program evaluation	96	44
Not sure	31	14
Total	**219**	**100**

There was substantial variation in how respondents in the same job classification perceived their performance appraisal as it related to program evaluation. In only a few cases all the respondents in the same job classification in each state answered similarly when asked if their performance was assessed in part on whether or not they conducted evaluation of their programs. This means that either some people don't know, people read the question differently, or there is variation in how different people are assessed within job classifications in the same state. A cross tabulation (Table 6-13) shows that nearly 61 percent of county/regional based respondents said that their performance was assessed in part on whether or not they conducted program evaluation, and nearly 30 said that it was not. About 10 percent was unsure. A smaller percentage of respondents with state-wide responsibilities (40 percent) said that their performance was assessed on whether they conducted program evaluation, and 20 percent was unsure. A majority (62%) of respondents agreed or strongly agreed that performance appraisals should include criteria related to program evaluation behaviour.

Table 6-13 Perceptions regarding performance appraisal
(Within job-classification cross tabulation)

The top percent indicates respondent ratio; the bottom number represents the number of respondents selecting the option.	**county/regional agent**	**state-wide specialist**	**program assistant**	**state forestry dept.**	**other**	**Total**
Believes performance assessed on whether they do evaluation	60.8% 45	39.6% 38	40.0% 2	8.1% 3	57.1% 4	42.0% 92
Believes it is not assessed on whether they do evaluation	29.7% 22	40.6% 39	40.0% 2	83.8% 31	28.6% 2	43.8% 96
Not sure	9.5% 7	19.8% 19	20.0% 1	8.1% 3	14.3% 1	14.2% 31
Total	100.0% 74	100.0% 96	100.0% 5	100.0% 37	100.0% 7	100.0% 219

Table 6-14 Performance assessments should include program evaluation criteria

	Frequency	**Percent**
Strongly agree	17	7.8
Agree	119	54.6
Neither agree nor disagree	55	25.2
Disagree	23	10.6
Strongly disagree	4	1.8
Total	218	100.0

6.3.6 Organizational commitment to evaluation

Sixty-six percent of respondents agreed or strongly agreed that program evaluation is a normal part of their organization. Eighty percent of respondents strongly disagreed with the statement that they are not expected to conduct evaluations of their programs.

Thirty-eight percent of respondents disagreed or strongly disagreed with the statement that their organization positively rewards Extension professionals for conducting program evaluations, and 37 percent was neutral. Only 24 percent agreed or strongly agreed.

Fifty-seven percent of respondents disagreed or strongly disagreed with the statement that their administrators are not committed to building evaluation capacity within the organization, and 26 percent was neutral. Thirty-seven percent of respondents agreed or strongly agreed that there are adequate resources within their organization to assist them with program evaluation. Just over one third believed that the resources within their organization to assist them with evaluation are inadequate.

Table 6-15 Perceived organizational commitment to program evaluation

The top percent indicates respondent ratio; the bottom number represents the number of respondents selecting the option.	**1 Strongly Agree**	**2 Agree**	**3 Neither Agree nor Disagree**	**4 Disagree**	**5 Strongly Disagree**	**Total**
1. Evaluation deeply rooted in administrative values	28% 61	43% 92	15% 33	13% 27	1% 3	100% 216
2. Not expected to conduct program evaluation	1% 3	9% 19	10% 21	38% 82	42% 91	100% 216
3. Organization rewards for conducting program evaluations	5% 11	19% 42	37% 80	28% 61	10% 22	100% 216
4. Administrators not committed to building evaluation capacity	4% 8	13% 29	26% 56	43% 93	14% 30	100% 216
5. Evaluation is normal part of organization's culture and practice	16% 34	50% 108	17% 36	16% 34	2% 5	100% 216
6. Adequate resources to assist me with program evaluation	6% 12	31% 66	31% 66	24% 51	10% 21	100% 216

Table 6-16 Perceived organizational commitment to program evaluation
(five point scale: 1 = strongly agree and 5= strongly disagree)

	Program evaluation deeply rooted in admin. values of my organization	**I am not expected to conduct evaluation of my programs**	**Organization positively rewards for conducting program evaluations**	**Administrators are not committed to building evaluation capacity**	**Evaluation is a normal part of my organization culture and practice**	**There are adequate resources to assist me with program evaluation**
Mean	2.2	4.1	3.2	3.5	2.4	3
Median	2	4	3	4	2	3
Mode	2	5	3	4	2	2
Std. Dev.	6.6	6.4	9.2	9.2	6.6	6.6
Std Error	.07	.07	.07	.07	.07	.07
Range	4	4	4	4	4	4
Minimum	1	1	1	1	1	1
Maximum	5	5	5	5	5	5

The summated score was obtained by averaging the average scores for each of six responses for each respondent. The distribution is normally distributed as shown in Figure 6-13 and Table 6-16. The mean, median, and mode were each 2.5. The maximum was 3.8 and the minimum was 1.3.

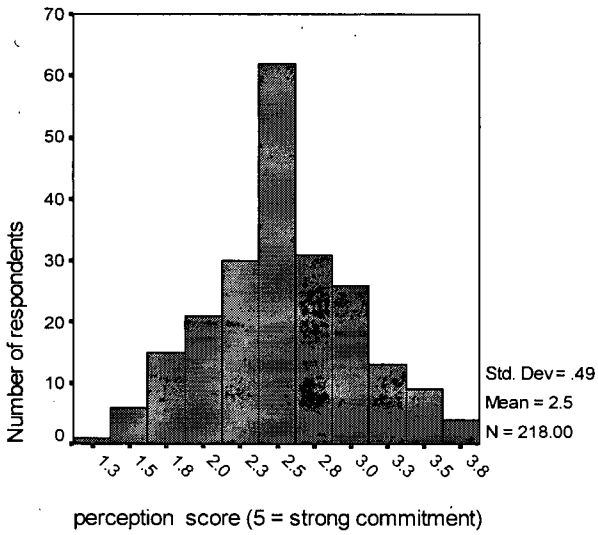

Figure 6-13 Perceived organizational commitment to program evaluation

6.3.7 Evaluation resources available to respondents

Respondents were asked to select all the program evaluation resources available to them. The resources most frequently selected were "colleagues who are willing to help me," "evaluation specialists I can call upon for help," and "in-service training in program evaluation." Six percent (12 respondents) said that no resources were available to them.

Table 6-17 Evaluation resources available

Resource	Frequency	Percent
Colleagues who are willing to help me with evaluation	151	69
Evaluation specialists I can call upon for help	117	53
In-service training in program evaluation	107	49
Publications on evaluation available to me	89	40
Not sure	24	11
Award programs for recognizing evaluation work	16	7
None	12	6
Other	8	4
Funds earmarked for evaluation	5	2

The "other" category included both comments and other resources:

- Budget for program evaluation in grants;
- Evaluation specialists, but they know more about agriculture evaluation than natural resources;
- Funds I earmark for evaluation, not administration;
- Graduate classes are available through department;
- In-service training is ineffective;

- There is a standard evaluation for all programs;
- There may be some others I do not know of; and
- Website with successful evaluation examples & types

6.3.8 Sources of external motivation

Respondents were asked to select among 11 sources of motivation to indicate ways that their administration motivates them to conduct program evaluation. Twenty-one percent (46 individuals) felt that their administration does not try to motivate them to conduct program evaluation. The most frequently selected source was "talking about evaluation frequently," followed by "providing evaluation specialists to help," and "providing evaluation training."

Table 6-18 Sources of external motivation for conducting program evaluation

Source	Frequency	Percent
Talking about evaluation frequently	91	42
Providing evaluation specialists to help	69	38
Providing evaluation training	63	29
Including evaluation plans into state Extension plans	61	28
Requiring evaluations as condition of promotion/tenure	61	28
Verbal "pats on the back" for conducting evaluations	60	27
Administration doesn't try to motivate	46	21
Using evaluation as criterion on performance appraisal	46	21
Administration doesn't try to motivate to do evaluation	46	21
Using my evaluation results in program decisions	45	21
Requiring evaluation as condition of continued funding	35	16
Being a role model by doing their own evaluations	16	7
Award programs to recognize outstanding evaluations	6	3

6.3.9 Greatest source of external motivation for evaluation

Respondents were asked to select an external motivation source that they felt would provide them with the greatest motivation to conduct program evaluation. There was a wide range in responses and no single external source was selected by greater than 16 percent of respondents. The most frequently selected motivation sources were "using evaluation results in program decisions," "requiring evaluation as a condition of promotion/tenure," and "using evaluation as a criterion on performance appraisals" at 16, 15, and 15 percent, respectively.

Table 6-19 Greatest source of external motivation

Source	Frequency	Percent
Using evaluation results in program decisions	33	16
Requiring evaluation as a condition of promotion/ tenure	32	15
Using evaluation as criterion on performance appraisals	32	15
Providing funds for evaluation	25	12
Providing evaluation specialists to help me	16	8
Requiring evaluation as condition of continued funding	15	7
Providing evaluation training	15	7
Other (Listed below)	14	7
Being a role model for evaluation	10	5
Award programs	9	4
Verbal "pats on the back"	4	2
Including evaluation plans into state plans	2	1
Talking about evaluation frequently	2	1
Subtotal	209	100
Missing	10	
Total	**219**	

The list of "others" includes:

- A check on the value of efforts to landowners;
- A combination of several of the above;
- Explaining how the information is used and why;
- Fund projects with good evaluations;
- Giving one generic evaluation form;
- (He does not) remind me;
- Improving program effectiveness and external grant;
- Pay me more;
- Providing evaluation instruments;
- Training in how evaluation can be used in program planning; and
- Money

Two respondents said that they were motivated by internal factors.

- Self - I want to know if my programs are effective; and
- Intrinsic motivation

6.3.10 Greatest barrier to evaluation

Respondents were asked to select from a list their greatest barrier to conducting program evaluation. Half of respondents selected lack of time, and nearly a quarter selected lack of skills as the greatest barrier. Only fourteen percent selected lack of incentive, but 13% listed "other" barriers, including a combination of the factors listed.

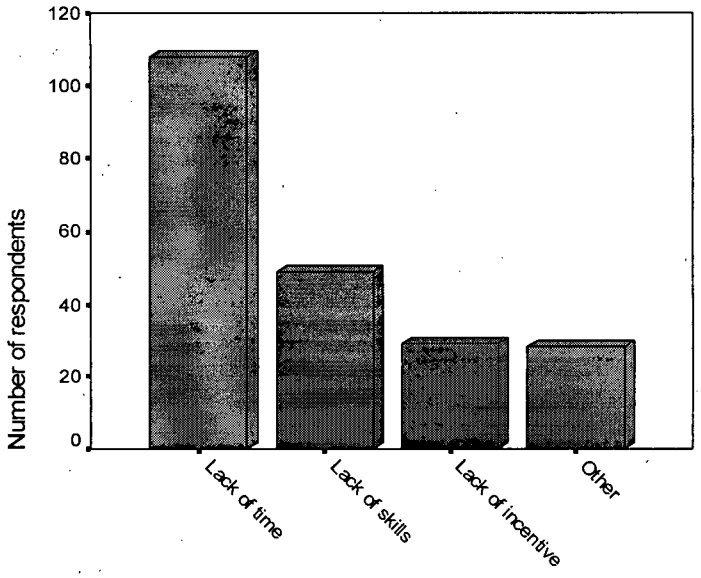

Figure 6-14 Greatest barriers to conducting program evaluation

Among those who selected "other," the following were included:

- Several respondents felt that the barriers were mixed:
  - A bit of each, I suppose;
  - All of the above;
  - All of the above; and
  - Both lack of time and skills equally.

- Others mentioned funding and resources:
  - Sufficient funding, including cost of my time;
  - Funding;
  - Funding!;
  - Primarily money and efficiency;
  - Lack of money;
  - Lack of resources including time and staff;
  - Lack of help from counties; and
  - Lack of support and assistance by specialists.

- Others mentioned barriers relating to evaluation methods:
  - Audience resistance to repeat evaluations;
  - Lack of interest; meeting participant revulsion;
  - Difficult program format/audience to evaluate;
  - Difficulty in measuring TRUE impact;
  - Difficulty to measure impact;
  - Disrupts my presentation & results are tainted;
  - Lack of response; and
  - Response rates (low)

- One person mentioned barriers in administration:
  - Two different administrators or levels.

- Several cited the lack in confidence in the value, utility, or priority of evaluation:
  - Not a priority;
  - Validity of the evaluation; and
  - Not sure evaluations are used in decision making

- Some respondents felt there were no barriers:
  - I evaluate every program;
  - There is no barrier;
  - There is NO barrier;
  - I usually do evaluations; and
  - None

6.4 Evaluation practices

6.4.1 Percent conducting evaluation

Seventy-nine percent of respondents (173 individuals) said that they had conducted some kind of program evaluation during the past year, while 21 percent (46 individuals) said that they had not.

Table 6-20 Conducted program evaluation in past 12 months

	Frequency	Percent
Conducted evaluation during past 12 months	173	79
Not conducted evaluation during past 12 months	46	21
Total	**219**	**100**

Of the individuals who conducted some kind of program evaluation during the past 12 months, 40 percent (68 individuals) said that they had conducted evaluation in addition to end-of event feedback or questionnaires and 60 percent (102 individuals) said that they had not.

Table 6-21 Conducted evaluation in addition to end of event questionnaires in past 12 months

	Frequency	Percent
Conducted evaluation in addition to end of event questionnaires	68	40
Not conducted evaluation in addition to end of event questionnaires	102	60
Total	**170**	**100**

Respondents were asked to describe the evaluations that they have conducted during the past year in addition to asking program participants to fill out questionnaires or provide verbal feedback at the end of an event. Responses ranged from needs assessments to measures of changes in KASAs, practice changes and changes in conditions. Both informal (such as "discussions with individual participants" and "input from advisory council") and formal methods (such as "mail surveys" and "documentation of forest management plans used") were used. A complete list of evaluation approaches used is found in Appendix D.

6.4.2 Level on Bennett's hierarchy

Respondents were asked how frequently they conducted various evaluation activities during the past 12 months. The scale ranged from 1 to 7 (with 1 as "never" and 7 as "always"). The activities were arranged hierarchically; it was expected that fewer respondents would measure long-term conditions than those who asked participants for their reaction to programs at the end of events.

Sixty percent of respondents said that they have asked participants for their reactions to a program at the end of an event "most of the time" or "always" in the past 12 months. Only two percent of respondents said that they have not done this in the past year.

Forty-four percent of respondents measure changes in knowledge, skills, attitudes, or aspirations "often, most of the time, or always" at the end of events. For other evaluation activities, the percent of respondents that do other activities "often, most of the time, or always" drops sharply. (Re-contacting = 15 percent of respondents, measuring behavioural change =16 percent of respondents, and measuring changes in long-term conditions = 10 percent of respondents).

Sixty-one percent of respondents said that they rarely or never measure long-term conditions. The percent jumps to 81 percent when "not often" is included in the count.

Table 6-22 Levels of evaluation conducted in past 12 months

The top percent indicates total respondent ratio; the bottom number represents the number of respondents selecting the option.	**1 Never**	**2 Rarely**	**3 Not often**	**4 About half the time**	**5 Often**	**6 Most of the time**	**7 Always**	**Total**
1. Asked participants for their reaction to a program at the end of an event.	2% 5	5% 11	5% 10	9% 20	19% 40	34% 73	26% 55	100% 214
2. Measured changes in knowledge, skills, attitudes, or aspirations at the end of an event.	9% 19	11% 24	20% 44	15% 33	15% 32	23% 50	6% 13	100% 214
3. Re-contacted program participants after program to assess changes in knowledge, skills, attitudes, or aspirations.	22% 48	23% 49	29% 61	11% 23	8% 18	7% 14	0% 1	100% 214
4. Measured changes in behaviours or practices of program participants as a result of programs.	22% 48	26% 55	28% 60	8% 17	10% 22	5% 10	1% 2	100% 214

Table 6-22, cont. Levels of evaluation conducted in past 12 months

The top percent indicates total respondent ratio; the bottom number represents the number of respondents selecting the option.	**1 Never**	**2 Rarely**	**3 Not often**	**4 About half the time**	**5 Often**	**6 Most of the time**	**7 Always**	**Total**
5. Measured changes in long-term conditions as a direct result of my programs.	32% 68	29% 63	20% 42	9% 19	7% 14	3% 7	0% 1	100% 214

Table 6-23 Levels of evaluation conducted in past 12 months (*seven point scale: 1= never and 7 = always)*

	Asked participants for reaction to program at end of events	**Measured changes in knowledge, skills, attitudes, or aspirations at end of event**	**Re-contacted participants to assess changes in knowledge, skills, attitudes, or aspirations**	**Measured changes in behaviours or practices**	**Measured changes in long-term conditions**
Mean	5.4	4.1	2.8	2.8	2.4
Median	6	4	3	3	2
Mode	6	6	3	3	1
Std. Deviation	1.5	1.7	1.5	1.5	1.4
Std. Error	.11	.12	.10	.10	.1
Range	6	6	6	6	6
Minimum	1	1	1	1	1
Maximum	7	7	7	7	7

6.4.3 Average of average scores on Bennett's Hierarchy

Average "program evaluation activity scores" for each individual were obtained by adding the values of the responses for each of the five levels for each respondent and dividing by number of responses. The table below shows the statistics of the averages (the average of the average). The mean, median, and mode of the evaluation activity average score was 3.8 on a seven-point scale, with standard deviation of 1.2 for 213 responses. The average score ranged widely, with the minimum at 1.0 and the maximum of 6.5. Figure 6-15 shows the range and frequency of average scores.

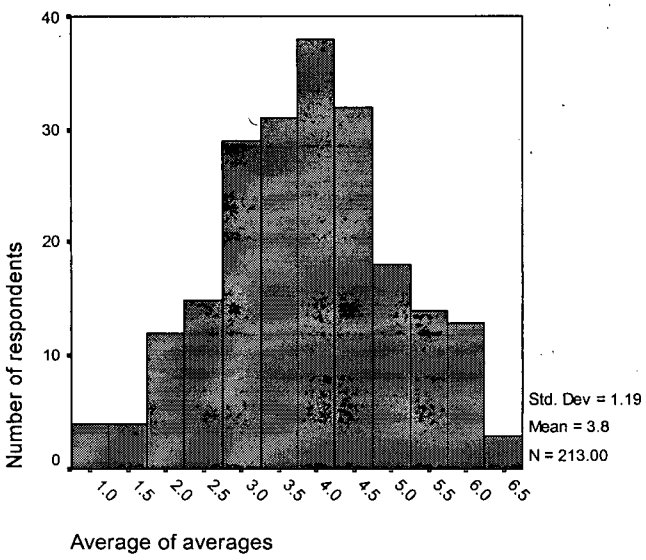

Figure 6-15 Average evaluation activity score (average of average)

6.4.4 Program evaluation level conducted most often

Respondents were asked to select the level on Bennett's Hierarchy at which they conducted evaluation most of the time. Just over 50 percent said that they ask participants for reactions at the end of events most of the time, and 32 percent measure changes in knowledge, skills, attitudes, and aspirations (KASAs) at the end of an event most of the time. Six percent re-contact participants after events most of the time. Eight percent measure behavioural changes most of the time, and .5 percent said that they measure long-term conditions most of the time. Six respondents did not answer the question.

Table 6-24 Levels evaluated most often

	Frequency	Percent
Don't evaluate programs	5	2
Ask participants for reactions	109	51
Measure KASAs at end of event	68	32
Re-contact participants after event	13	6
Measure behavioural changes	17	8
Measure long-term conditions	1	1
Sub-total	213	100
Missing	6	
Total	219	

6.4.5 Information sought in past year

Thirty-two percent of respondents (71 respondents of 219) said that they did not seek information on program evaluation in the past year. Sixty-eight respondents (31%) relied on expertise of others on their team. Sixty-seven respondents (30%) asked colleagues in their offices for help. Fifty-eight respondents (26%) used reference materials that they own and

fifty-seven respondents (26%) contacted Extension evaluation specialists on campus. The 13 respondents who selected "other" provided the following sources:

- Assistance by District Agents;
- Attended evaluation workshop;
- Examples from other agencies, Extension programs;
- EXTENSIVELY reading, information system/knowledge management;
- I talk to an evaluation and assessment professor;
- Looked at past surveys;
- My regional chair;
- On line contacts;
- Outside statistical/survey training;
- Participated in an evaluation workshop;
- Personal satisfaction of achieving goals;
- Professional journal articles; and
- Work with colleagues in other departments.

Table 6-25 Source of information about evaluation sought in past year

	Frequency	Percent
Didn't seek information on evaluation	71	32
Relied on expertise of others on my teams	68	31
Asked colleagues in my office for help	67	30
Reference materials that I own	58	26
Evaluation specialists on campus	57	26
Reference materials provided for me	47	21
Program evaluation websites	36	16
Reference materials that I borrowed	31	14
Others	13	7

6.5 Desired evaluation skill levels

Respondents were asked to select between "I know a lot," "I know a bit," and "I know enough now" in 11 program evaluation areas. No program evaluation skill had over 43 percent of respondents saying that they "know enough now." Only 17 percent of respondents said that they knew enough about developing survey instruments. Forty-three percent said that they knew enough about preparing evaluation reports.

Table 6-26 is sorted by the category "I need a lot more skill" in descending order. It shows that the highest percentage of respondents need more skill in choosing sampling techniques, developing survey instruments, using evaluation reports, and conducting focus groups.

Table 6-26 Desired skill level

The top percent indicates total respondent ratio; the bottom number represents the number of respondents selecting the option	**I know enough now**	**I need a bit more skill**	**I need a lot more skill**	**Total**
Choosing sampling techniques	16% 35	47% 102	36% 78	100% 215
Developing survey instruments	17% 36	51% 108	31% 66	100% 210
Using evaluation results	23% 50	50% 107	27% 58	100% 215
Focus group interviews	27% 58	46% 98	27% 57	100% 213
Conducting needs assessments	32% 68	44% 95	24% 52	100% 215
Developing evaluation plans	32% 68	44% 95	24% 52	100% 215
Selecting evaluation methods	40% 85	40% 86	20% 44	100% 215
Analyzing evaluation data	26% 56	58% 124	16% 35	100% 215
Testing a survey instrument	32% 70	51% 111	16% 35	100% 216
Preparing evaluation reports	43% 92	41% 88	16% 35	100% 215
Writing measurable objectives	40% 85	47% 100	14% 30	100% 215

The following three tables summarize the frequencies and percentages of the top six evaluation areas in each of three categories, "need a lot more skill," "need a bit more skill" and "know enough now."

Table 6-27 Topics that "need a lot more skill" *(most frequent six)*

	Number of respondents	**Percent of respondents**
Choosing sampling techniques	78	36
Developing a survey instrument	66	31
Using evaluation results	58	27
Conducting focus group interviews	57	27
Conducting needs assessments	52	24
Developing evaluation plans	52	24

Table 6-28 Topics that "need a bit more skill" *(most frequent six)*

	Number of respondents	**Percent of respondents**
Analysing evaluation data	124	58
Developing survey instruments	108	51
Testing a survey instrument	111	51
Choosing sampling techniques	102	47
Writing measurable objectives	100	47
Focus group interviews	98	46

Table 6-29 Topics about which "know enough now" *(most frequent six)*

	Number of respondents	Percent of respondents
Preparing evaluation reports	92	43
Writing measurable objectives	85	40
Selecting evaluation methods	85	40
Testing a survey instrument	32	70
Developing evaluation plans	68	32
Conducting needs assessments	68	32

6.6 Review of Extension performance appraisal forms

One of the hypotheses of this study is that higher levels of evaluation are conducted by NREPs who believe that their own job performance is appraised on the basis of their program evaluation behaviour. I tested this hypothesis through a survey question that directly asked respondents if they thought their performance was appraised on their evaluation behaviour. I also collected copies of performance appraisal forms from 39 states over a six-month period in 2002 and reviewed them to determine whether or not "conducting evaluation of Extension programs" or related verbiage actually appeared among the performance criteria. I assumed that the presence of criteria relating to program evaluation behaviour on the forms was an indicator of organizational commitment to evaluation.

I found that performance appraisal forms clearly include criteria related to program evaluations in some states, while expectations seem vaguer in other states. Of the 39 states that provided performance appraisal forms, 48 percent (19 states) included vague or no reference to program evaluation. In other states, there were some references to evaluation such as "demonstrating impacts." Some referred generally to "scholarly accomplishment," which typically refers activities such as presentations at professional conferences and involvement in research (but could potentially include evaluation). In one state, the same evaluation form was used for classified staff, unclassified staff, and faculty and no reference to program evaluation was mentioned. Frequently, the performance appraisal for Extension personnel was conducted in the same way as all other faculty (such as a dossier every three years outlining courses taught, publications, grants received, professional service, and presentations conducted).

As an example of a state where program evaluation is clearly referenced in the performance appraisal criteria, the Oregon Extension Periodic Faculty Review and Professional Development Plan criteria includes the following references to program evaluation (see Appendix F):

Teaching, Facilitation, and Other Assignments:

- "identifies needs";
- "identifies measurable goals and objectives";
- "evaluates educational programs to determine if goals and objectives are met and to document long-term outcomes."

Program Delivery:

- "designs educational programs to support behavior change among participants, documented through impact assessment";
- "clearly documents the impact of educational activities on economic well-being, environmental quality, and/or quality of life of target audience";

Scholarship and Creative Activity:

- "participates in or initiates leadership in intellectual work, or other creative activity, including evaluative processes."

Organizational Accountability and Maintenance:

- "documents measurable program outcomes on at least one priority education event annually."

Another example of a state where evaluation is clearly indicated in the performance appraisal criteria was North Carolina's "Extension Management and Performance System" (E-MAPS) that included references to evaluation in most categories. The criteria related to the highest rating possible (in other words, if the NREP conducts evaluation, they would receive the highest rating for a given category).

Only a few other states such as Alabama, Tennessee, Kentucky, Mississippi, and Kansas had criteria as specific or included reference to evaluation as frequently as Oregon's and North Carolina's in 2002. Assuming that the forms are used and the criteria clearly communicated, these faculty members clearly appear to be expected to conduct program evaluations.

6.7 Factors associated with level of Program Evaluation conducted

6.7.1 Cluster analysis

Cluster analysis is a multivariate technique used to partition respondents into homogeneous groups based on one or more of their characteristics (Hair et al. 1998). The technique is an exploratory approach to classifying respondents so that each respondent in a group is similar to the others in the same group with respect to one or more key characteristics selected by the researcher. This allows the researcher to look for similarities and differences between groups based on how they group around key variables. I chose this technique as an exploratory exercise to see if it revealed trends among my variable of interest, "level of program evaluation conducted most of the time" that may not have been revealed through other analyses. I conducted the analysis in SPSS using five subquestions of question 15 (level of program evaluation conducted most of the time) as the grouping variables and then looked at the characteristics of the respondents that fell into each of the groups.

I conducted the analysis using the K-means (non-hierarchical) clustering method to find underlying groupings. A non-hierarchical method was selected because of its relative advantage with continuous variables and larger samples, compared with hierarchical approaches.

The non-hierarchical method uses an iterative (trial and error) process of determining number of groups. I first tried four groups (clusters) but then settled on three groups because the three-group option showed the most homogeneity within groups and the most heterogeneity between groups.

Hair et al. (1998) discuss two assumptions of cluster analysis: representativeness of sample and freedom from multicollinearity. The representativeness of the sample was addressed through the non-response survey. The multicollinarity assumption was tested by running chi-square statistics for each pair to test associations. Some collinearity was found, however, the cluster analysis is not being used to draw inferences but is exploratory in nature; often multicollinearity is tolerated when the objective is exploratory (R. Kozak, Associate Professor, Faculty of Forestry, University of British Columbia, pers. comm., July 2004).

Table 6-30 shows the result of the iterative selection process that determined that three groups was the best choice of number of groups. The numbers in the tables below refer to the means for each variable (on a scale of 1-7 based on survey question 15). Table 6-31 shows the number of cases per cluster in the final three-group selection. After I determined that three clusters was the best choice of number of groups, I named them "Keeners," "Middle roaders," and "Foot draggers," which characterized their approach to program evaluation. Then I ran frequency statistics for each cluster by "selecting cases" in SPSS as shown in Table 6-32.

Table 6-30 Final Cluster Centers

Final cluster centers Levels of program evaluation conducted most of the time	**Cluster 1 "Keeners"**	**Cluster 2 "Middle roaders"**	**Cluster 3 "Foot-draggers"**
Asked participants' reactions at end of event	6	6	4
Measured KASA (knowledge, attitudes, skills, and aspirations) changes at end of event	6	5	2
Re-contacted participants to assess KASA changes	5	3	2
Measured behavioural changes after event	5	3	2
Measured long term conditions after event	4	2	2

Table 6-31 Number of Cases per cluster

Cluster 1 "Keeners"	47
Cluster 2 "Middle roaders"	90
Cluster 3 "Foot-draggers"	74

Table 6-32 below summarizes the results of the cluster analysis. The % refers to the percentage of respondents who fell into that cluster. The respondents naturally fell into these three clusters and characteristics of most variables followed a distinct pattern.

Table 6-32 Results of Cluster Analysis for three groups

Variable	Cluster 1 "Keeners"	Cluster 2 "middle roaders"	Cluster 3 "Foot-draggers"
Number of respondents per cluster	47	90	74
Have conducted any program evaluation in past year	93%	90%	59%
Have conducted evaluation beyond "end of event" in the past year	55%	31%	14%
Gut reaction to program evaluation	Love it 17% Don't mind it 68% Prefer to ignore 13% Dread it 0% Rarely think of it 0%	Love it 6.7% Don't mind it 64% Prefer to ignore 26% Dread it 0% Rarely think of it 0%	Love it 5% Don't mind it 47% Prefer to ignore 37% Dread it 4% Rarely think of it 5%
Average percent motivation for conducting evaluation derived from external sources	44% SD= 24%	55% SD=25%	61% SD=28%
Average attitude score *(scale 1-5 where 1 is "strongly agree" with positive statements about attitude toward program evaluation)*	Mean = 2.23 a Median =2.14 SD=.61	Mean = 2.44 b[1] Median = 2.43 SD=.58	Mean = 2.72 b Median = 2.71 SD=.52
Average perceived organizational commitment score *(scale 1-5 where 1 is strongly agree with positive statements about organizational commitment to program evaluation)*	Mean = 2.38 a SD= .68	Mean = 2.40 b SD=.67	Mean =2.80 b SD=.68
Average age	47 years SD=9 years	44 years SD=9 years	44 years SD= 11 years
Average years experience in Extension	15 years SD= 9 years	10 years SD= 8 years	13 years SD=11

[1] Means within a row followed by the same letter are not significantly different. Scheffe Post Hoc test, $p<.05$

Table 6-32, cont. Results of Cluster Analysis for three groups

Variable	Cluster 1 "Keeners"	Cluster 2 "middle roaders"	Cluster 3 "Foot-draggers"
Percent of respondent salary that is grant-funded (percent of cluster)	None of salary 45% 1-50% grant funded 26% 51-99% grant funded 6% All salary grant funded 21.3%	None of salary 56% 1-50% grant funded 22% 51-99% grant funded 13% All salary grant funded 9%	None of salary 55% 1-50% grant funded 22% 51-99% grant funded 10% All salary grant funded 12%
Position classification	Tenure track 33% non-tenure-track 31% professional staff 34% other 3%	Tenure track 33% non-tenure-track 27% professional staff 37% other 0/0%	Tenure track 31% non-tenure-track 17% professional staff 47% not sure 1% other 1%
Have achieved tenure (% of tenure-track respondents)	36.0%	22.0%	28.4%
Highest education level	Bachelor 10% Master 53% PhD 36%	Bachelor 16% Master 56% PhD 29%	Bachelor 29% Master 40% PhD 29%
Believes their own performance appraisal includes criteria relating to program evaluation	55.0%	46.7%	28%
Confidence in conducting evaluation	Very confident 31% Moderately confident 51% Neutral 0% Moderately unconfident 0% Very unconfident 2%	Very confident 6.7% Moderately confident 50% Neutral 28.9% Moderately unconfident 13.3% Very unconfident 0%	Very confident 5% Moderately confident 26% Neutral 35% Moderately unconfident 28% Very unconfident 5%

Table 6-32, cont. Results of Cluster Analysis for three groups

Variable	Cluster 1 "Keeners"	Cluster 2 "middle roaders"	Cluster 3 "Foot-draggers"
Five most frequently mentioned sources of motivation for conducting evaluation	1) using evaluation results in decision making (26%) 2) providing funds for evaluation (15%) 3) requiring evaluation as a condition of promotion and tenure (13%) 4) including evaluation as criterion on performance evaluations (11%) 5) providing evaluation specialists (89%)	1) requiring evaluation as a condition of promotion and tenure (19%) 2) using evaluation results in decision making (13%) 3) including evaluation as criterion on performance evaluations (13%) 4) providing funds for evaluation (12%) 5) requiring evaluation as a condition of continued funding (9%)	1) including evaluation as criterion on performance evaluations (18%) 2) using evaluation results in decision making (12%) 3) requiring evaluation as a condition of promotion and tenure (12%) 4) providing evaluation training (11%) 5) providing funds for evaluation (10%)
Evaluation specialists available to help	72%	54%	41%

With the exception of a few variables, "keeners" scored highest on most positive aspects of evaluation, "foot-draggers" scored the lowest, and "middle roaders" were in the middle. The "keeners" (cluster 1) and "middle-roaders" (cluster 2) had much higher percentages of respondents who had conducted evaluation in the past 12 months compared with "foot-draggers" (cluster 1). Fifty-five percent of respondents in cluster 1 had conducted evaluations beyond "end of event" evaluations, whereas only 31 percent of "middle roaders" and only 14 percent of "foot-draggers" did.

A much higher percentage of "keeners" said that they love doing evaluation than the other clusters, and a far smaller percentage of "keeners" preferred to ignore evaluation than the other two groups. "Keeners" said that a smaller percentage of their motivation to conduct program evaluation was due to external motivating factors, implying that a greater percentage of their motivation was derived intrinsically than the other two groups.

A higher percentage of "keeners" had positive attitudes towards evaluation, perceived that their organizations were committed to evaluation, and believed their own performance appraisal included criteria relating to program evaluation than the other two groups. A much higher percentage of "foot-draggers" had bachelor's degrees as their highest degree and a higher percentage of "keeners" had slightly more PhDs than the other two groups. "Keeners"

felt much more confident in conducting program evaluations than either "middle roaders" or "foot-draggers." "Keeners" were slightly older and more experienced, but not substantially so. Position classification (tenure-track vs. non-tenure track), age, and tenure status did not follow patterns.

"Keeners" listed "using evaluations in program decisions" more frequently as their top external motivating factor for conducting program evaluation than the other two groups. All respondents who said that they dread doing evaluation fell into the laggard group. "Foot-draggers" have a higher percentage of motivation that is externally derived than the other two groups (61 percent compared with 44 percent for "keeners" and 55 percent for "middle-roaders)." "Middle roaders" and "foot-draggers" tended to rely more on external motivators such as promotion and tenure or using evaluation as a criterion on performance evaluations as their top motivator.

"Keeners" had a higher percentage of respondents who perceived that they had access to evaluation specialists than the other two groups. Seventy-two percent of "keeners" believed they had access to evaluation specialists, 54 percent of "middle roaders," and only 40 percent of "foot-draggers" believed they had access to evaluation specialists to help them.

A smaller percentage of "foot-draggers" believed that their performance is evaluated on the basis of conducting program evaluation. Only 40 percent of "foot-draggers" said that they have evaluation specialists to help, compared with 72 percent of "keeners" who said that they have evaluation specialists available to help.

While it is tempting to draw predictive or generalized statements based on the cluster analysis, one must realize that the purpose of cluster analysis is to group, not to generalize or make inferences. The other two analyses, ANOVA and Multiple linear regression are more statistically reliable and are intended for drawing conclusions about relationships between factors.

"Foot-draggers" had the lowest overall average attitude score. Forty-two percent of the cluster was less than 100 percent core funded. A much higher proportion consisted of professional staff but members of this cluster covered all job classifications. They were almost equally as likely to have bachelor's as have a PhD as highest degree and a much higher proportion were moderately unconfident and very unconfident, and felt that performance appraisals and promotion and tenure were the greatest motivators for them for conducting evaluation.

One-way ANOVAs were conducted to see if there were significant differences between clusters with regards to perceived organizational commitment, attitude towards evaluation, and age. The ANOVA was conducted using the three clusters as grouping variables. Since age and years of experience were highly correlated, age was included in the test but not years of experience. Levene's test showed that the variables met the assumption of homogeneity of variance (Perceived organizational commitment F= .197 p= .821, Attitude F= 1.012, p=.365, and Age, $F = 2.211$, $p = .112$)

The ANOVA revealed that there were significant differences between means of perceived organizational commitment and attitude, but not age. Scheffe's Post Hoc test showed that for perceived organizational commitment, there were significant differences between foot-draggers and "keeners" (groups 1 and 3) and between middle-roaders and "keeners" (groups 2 and 3). For attitude, significant differences existed between middle-roaders and "keeners" (groups 2 and 3) and foot-draggers and "keeners" (groups 1 and 3).

Table 6-33 ANOVA- Perceived organizational commitment, attitude, and age

		Sum of Squares	df	Mean Square	*F statistic*	*P value*
Perceived organizational commitment	Between Groups	8.05	2	4.03	8.7	.000
	Within Groups	96.15	208	.462		
	Total	104.20	210			
Attitude toward program evaluation	Between Groups	6.44	2	3.2	9.95	.000
	Within Groups	67.34	208	.324		
	Total	73.79	210			
Age	Between Groups	435.50	2	217.75	2.36	.097
	Within Groups	18512	201	92.10		
	Total	18948	203			

6.7.2 Cross tabulations/correlations

Cross tabulations show the relationships between categorical independent variables and the dependent categorical variable, "way you have conducted program evaluation most of the time in the past 12 months." I ran cross tabulations on four categorical independent variables related to my hypotheses (funding source, position classification, tenure status, and whether or not supervisors assess performance on basis of program evaluation) as well as two other variables, (job classification to see how county/region-based and campus-based respondent compared and confidence level to see if confidence levels made a difference).

Table 6-34 below shows a summary of the variables examined, and results of a chi-square test for correlations. Pearson's chi square statistics showed that most categorical variables showed no significant relationship with the variable, "Ways evaluated most often in the past 12 months" except confidence level and job classification.

Table 6-34 Tests for associations between "Ways evaluated most often in the past 12 months" and other categorical variables

Variable	*Chi-square value*	*P value*
Variables related to hypotheses		
1) Funding source (categories)	15.65	.075
2) position classification (tenure/non-tenure)	9.35	.673
3) Tenure status	7.31	.605
4) Whether supervisor assesses performance on basis of program evaluation	9.51	.147
Other variables examined		
1) Confidence level	25.6	.012
2) Job classification (county, state, professional staff, program assistant)	31.22	.002

The alpha level was set at the .05 level for all significance tests. Df= 5

The following tables are details of the cross tabulations for each variable mentioned in Table 6-34 above.

6.7.2.1 Funding source

Funding source was recoded from a continuous into a categorical variable because it was not normally distributed. The cross tabulation and chi-square test shows that there is not a pattern between funding source and evaluation behaviour. Surprisingly, core funded NREPs conducted as many higher level evaluations as grant-funded NREPs.

Table 6-35 Funding source * Ways you evaluate most often Cross tabulation

The top percent indicates respondent ratio; the bottom number represents the number of respondents selecting the option.	ask for reaction	measure KASA at end	re-contact participants	behaviours and long term conditions	Total
0 % grant funded	55.5% 61	34.5% 38	3.6% 4	6.4% 7	100.0% 110
1-50 % grant funded	56.0% 28	18.0% 9	12.0% 6	14.0% 7	100.0% 50
51-99 % grant funded	42.9% 9	42.9% 9		14.3% 3	100.0% 21
100 % grant funded	44.0% 11	40.0% 10	12.0% 3	4.0% 1	100.0% 25

6.7.2.2 Position classification

The same percentage of tenure-track and non-tenure track faculty selected behaviours and long-term conditions as their most frequent evaluation method (6%), and almost the same percentage re-contacted participants after events to assess KASAs (7.2% and 6% respectively). A higher percentage (31% verses 40%) of non-tenure track faculty members measure KASAs as their most frequent method than tenure track faculty. About the same percentage of professional staff selected "ask for reactions" as the other two categories, but a higher percentage of professional staff measure behavioural and long-term changes than the other two categories and a much smaller percentage of professional staff measure KASAs than the other two categories. Of those who selected "measure behavioural or long-term changes" most of the time, 61 percent were professional staff. This seems to indicate that whether or not a NREP position is tenure-track does not influence the level of program evaluation the NREP conducts, unless they are professional staff, in which case a higher percentage are more likely to conduct higher levels of evaluation.

Table 6-36 Position classification * Ways you evaluate most often Cross tabulation

The top percent indicates respondent ratio; the bottom number represents the number of respondents selecting the option.	**ask for reaction**	**measure KASA at end**	**re-contact participants**	**behaviours and long term conditions**	**Total**
Tenure track faculty	56.5% 39	30.4% 21	7.2% 5	5.8% 4	100.0% 69
Non-tenure track faculty	48.0% 24	40.0% 20	6.0% 3	6.0% 3	100.0% 50
Professional staff	53.8% 43	26.3% 21	6.3% 5	13.8% 11	100.0% 80
Not sure	33.3% 1	66.7% 2			100.0% 3
Other, please specify	33.3% 1	66.7% 2			100.0% 3

6.7.2.3 Tenure status

Among tenure-track respondents, both tenured and not-yet-tenured faculty ask for reactions as their most frequent evaluation method about equally (55% for tenured and 52% for not-yet-tenured). A higher percentage (43%) of those who have not yet achieved tenure measure KASAs at the end of events than tenured faculty (26%), but a smaller percentage of not-yet-tenured faculty (4.3% verses 10.3%) re-contact participants. No respondents who have not yet achieved tenure selected "measure behaviours or long-term conditions." There does not seem to be a strong pattern between tenured and not-yet-tenured regarding their level of program evaluation. A higher percentage of tenured faculty than not-yet-tenured faculty measure behavioural and long-term changes. The process of achieving tenure may not be a strong incentive for higher level program evaluations.

Table 6-37 Tenure status * Ways you evaluate most often Cross tabulation

The top percent indicates respondent ratio; the bottom number represents the number of respondents selecting the option.	**ask for reaction**	**measure KASA at end**	**re-contact participants**	**behaviours and long term conditions**	**Total**
Not tenure-track faculty	50.8% 64	34.1% 43	4.8% 6	10.3% 13	100.0% 126
Have achieved tenure	55.2% 32	25.9% 15	10.3% 6	8.6% 5	100.0% 58
Have not achieved tenure	52.2% 12	43.5% 10	4.3% 1		100.0% 23
Not sure	100.0% 1				100.0% 1

6.7.2.4 Whether or not supervisors assess performance on basis of program evaluation

For those who measure reactions as their most frequent evaluation method, a higher percentage (48 %) did not believe their performance is assessed on the basis of program evaluation than those who believed their performance is measured on basis of program evaluation activities (37 %). Sixteen percent were unsure. Contrastingly, a much higher percentage of respondents who measure KASAs as their most frequent evaluation method said that their performance is assessed on their program evaluation activities (57 percent). The percentage of respondents that believed their performance is assessed on the basis of program evaluation declined when considering higher levels of Bennett's Hierarchy.

There is some variability among respondents from the same state and job classification regarding their perception of whether or not their performance is assessed on the basis of their program evaluation activities. In 16 states, respondents in the same job classification differed in their response (i.e. at least one person said their performance was assessed on basis of their program evaluation activities, while at least one person said it was not). This occurred with respondents from Arizona, Florida, Colorado, Idaho, Indiana, Kentucky, Louisiana, Michigan, Minnesota, New York, Oregon, Pennsylvania, South Carolina, Tennessee, Utah, Washington, and Wisconsin. Respondents from nineteen states said that they were unsure. While there is a possibility that some respondents in the same job classification work under different evaluation criteria, and some states had only a few respondents (which increases the possibility of error), some respondents with the same

performance appraisal criteria appeared to have different perceptions about program evaluation expectations in at least some states. Disagreement in perceptions among people from the same state and job classification and the high number of "not sure" responses could be an indicator of weak communications regarding expectations and criteria for performance evaluations from administrators.

Table 6-38 Supervisors assess performance in part on whether or not you conduct program evaluation of programs * ways evaluated most often in the past 12 months Cross tabulation

The top percent indicates respondent ratio; the bottom number represents the number of respondents selecting the option.	**Ask for reaction**	**measure KASA at end**	**re-contact participants**	**behaviours and long term conditions**	**Total**
Yes	40 44.0%	39 42.9%	4 4.4%	8 8.8%	91 100.0%
No	52 58.4%	24 27.0%	6 6.7%	7 7.9%	89 100.0%
Not sure	17 60.7%	5 17.9%	3 10.7%	3 10.7%	28 100.0%

6.7.2.5 Confidence level

Sixty-seven percent of those who measure behavioural or long-term changes indicated that they felt confident or moderately confident in conducting program evaluation, and none selected very unconfident. Twenty-four percent of those who ask for reactions as their most frequent evaluation method are moderately unconfident or very unconfident, and 80 percent of moderately unconfident respondents indicated that "asking for reactions" is their most frequent evaluation method. This shows that there is a trend linking confidence levels in conducting program evaluation and levels of conducting program evaluation. Higher levels of program evaluation seem to be linked with higher levels of confidence in conducting program evaluation.

Table 6-39 Confidence level in conducting program evaluations * ways evaluated most often in the past 12 months Cross tabulation

The top percent indicates respondent ratio; the bottom number represents the number of respondents selecting the option.	**ask for reaction**	**measure KASA at end**	**re-contact participants**	**behaviours and long term conditions**	**Total**
Very confident	20.0% 5	44.0% 11	12.0% 3	24.0%6	100.0% 25
Neither confident nor unconfident	51.7% 30	34.5% 20	6.9% 4	6.9% 4	100.0% 58
Moderately unconfident	80.0% 24	13.3% 4		6.7% 2	100.0% 30
Very unconfident	50.0% 2	50.0% 2			100.0% 4

6.7.2.6 Job Classification

A greater proportion of state-level NREPs (9.8%) assessed behavioural and long-term impacts than county-level NREPs (2.7%) as the way they evaluated most often in the past 12 months. However, a much higher proportion of county-level NREPs (51.4%) assessed KASA changes at the end of events than state-level NREPs (26.1%). Sixty-four percent of state forestry personnel assessed reactions at the end of events as the way in which they evaluated most often in the past 12 months. Surprisingly, 22 percent of state forestry employees said that they measure behaviours and long-term conditions most of the time.

Table 6-40 Job classification * ways evaluated most often in the past 12 months Cross tabulation

The top percent indicates respondent ratio; the bottom number represents the number of respondents selecting the option.	**Ask for reaction**	**measure KASA at end**	**re-contact participants**	**behaviours and long term conditions**	**Total**
Extension agent with county/regional responsibilities	29 39.2%	38 51.4%	5 6.8%	2 2.7%	74 100.0%
Extension specialist with State-wide responsibilities	54 58.7%	24 26.1%	5 5.4%	9 9.8%	92 100.0%
State forestry dept. employee	21 63.6%	3 9.1%	2 6.1%	7 21.2%	33 100.0%
Other	5 55.0%	3 34.0%	1 11.0%		9 100.0%

6.7.3 Multiple linear regression

I used a multiple linear regression to test the contributions of several independent variables in predicting the dependent variable, program evaluation behaviour (summated survey question 15). The independent variables were:

- attitude (summated score);
- perceived organizational commitment (summated score);
- age;
- years of experience in Extension;
- believes performance is appraised on basis of program evaluation (dummy variables);
- position classification (dummy variables); and
- tenure status (dummy variables).

I used four continuous variables and eight dummy variables (from three survey questions with nominal variables) as the predictor variables for the procedure. (I recoded three survey questions that had a total of eight selections to create eight dummy variables that I included in the regression).

There are four assumptions of multiple regression that are considered critical to test (Osborne and Waters 2002). They are:

- Variables are normally distributed. Through visual inspection of the histograms, I observed that most of the variables were normally distributed. 'Years of experience'

was not normally distributed but I conducted a data transformation using square root to compensate and ran the regression with the transformed variable.

- Linear relationship between independent and dependent variables. I examined scatterplots of residuals that indicated linear relationships existed.
- Variables are measured reliably. I used Cronbach's alpha to ensure reliability of the index scale questions and eliminated sub-questions to increase reliability. Osborne and Waters (2002) consider .7 to .8 as acceptable for Cronbach alpha (Cronbach alpha was .77 and .78 in this study for the two index scale variables)
- Constant variance among the independent variables. I tested for homoscedasticity by examining a plot of the standardized residuals (errors) to ensure that the residuals were randomly scattered.

Table 6-41 Correlation Matrix

		Age	Perceived org. Commitment	Supervisor's assess on evaluation behaviour	Position classification	Atttitude
Age	Pearson's correlation	1.0				
	Sig.	---				
Perceived org. commitment	Pearson's correlation	-.001	1.0			
	Sig.	.495	---			
Supervisor's assess on evaluation behaviour	Pearson's correlation	.026	-.507	1.0		
	Sig.	.356	.000*	---		
Position classification	Pearson's correlation	.282	.211	.175	1.0	
	Sig.	.000*	.001*	.007*	---	
Atttitude	Pearson's correlation	.078	.099	.036	-.001	1.0
	Sig.	.136	.083	.305	.496	---

I ran the regression using a backward stepwise elimination method that created eight models. The eighth model had an r^2 value of .223, meaning that only 22 percent of the variability in evaluation behaviour could be explained by the independent variables. The regression showed that the remaining five independent variables in the model had an *F value* of 11.03 at $p = .000$. The variables "tenure status" and "years of experience" dropped out of the model. Table 6-42 below shows the results of the backward elimination method which shows the eighth model with the r^2 value of .223.

The correlation matrix above shows that there were significant correlations among three variables: position classification, supervisors assess performance on basis of evaluation

behaviour, and perceived organizational commitment. This weakens the confidence in those variables as influencers of program evaluation behaviour because of the correlations.

Table 6-42 Regression model summary

Model	R	R^2	Adjusted R^2	Std. Error	R^2 Change	F Change	df1	df2	Sig. F Change
1	.477	.227	.177	1.0785	.227	4.531	12	185	.000
2	.477	.227	.181	1.0756	.000	.016	1	187	.900
3	.476	.227	.186	1.0728	.000	.015	1	188	.901
4	.476	.227	.190	1.0700	.000	.022	1	189	.883
5	.476	.227	.194	1.0673	.000	.032	1	190	.858
6	.476	.226	.198	1.0647	.000	.098	1	191	.755
7	.475	.226	.202	1.0623	-.001	.138	1	192	.711
8	.472	.223	.203	1.0614	-.003	.665	1	193	.416

Table 6-43 shows the remaining independent variables in the model and their coefficients. The collinearity diagnostics show that the variance inflation factor (VIF) and tolerance are within acceptable limits. (A VIF greater than 10 and a tolerance less than .1 indicate a serious problem with collinearity (Fields 2000). The *t statistic* indicates whether the variable is making a significant contribution to the model. The dummy variables associated with belief that supervisors assess their performance on the basis of program evaluation and position classification are significant in the model but are weaker predictors of program evaluation behaviour because of their low *t statistic*.

Table 6-43 Independent variables remaining in the multiple linear regression model

Independent variables	Variable type	Unstandardized coefficient	*t statistic*	*P value*
Attitude summated score	continuous	-.682	-5.34	.000
Perceived organizational commitment summated score	continuous	-.322	-2.56	.011
Age	continuous	.016	1.99	.048
Supervisor assesses performance on basis of program evaluation (dummy variable)	dummy	-.356	-.20	.045
Position classification (tenure-track)	dummy	-.377	-.217	.031

Table 6-44 Collinearity diagnostics of regression model

	Tolerance	VIF
Attitude summated score	.984	1.017
Tenure track faculty (dummy variable)	.868	1.152
Supervisor does not assess performance on basis of program evaluation (dummy variable)	.735	1.361
Age	.908	1.101
Perceived organizational commitment	.720	1.388

The regression reveals that five variables have linear relationships with evaluation behaviour, but only a small portion of the variance in evaluation behaviour can be explained by the variables. It shows that attitude, perceived organizational commitment to evaluation, age, whether or not NREPs' performance is assessed on the basis of evaluation behaviour, and whether or not they have a tenure-track position do slightly influence evaluation behaviour. Attitude towards evaluation is the strongest contributor to the model ($p < .001$), and age is the weakest contributor ($p = .048$).

I also ran a regression using the dependent variable and the seven individual items (sub-questions) that were used as indicators for attitude. There were several variables that showed linear relationships but the test was abandoned because it violated the assumption of no multicollinearity. Given that the items were used as an index and Cronbach's alpha showed high intercorrelation among the items, I assumed that a regression using individual attitude variables would give erroneous results.

Overall, the regression procedure did not show strong linear relationships between evaluation behaviour and the other continuous variables; the low r^2 value (.223) indicates that very little of the variation in the dependent variable can be explained by the combination of independent variables. Further, the correlations among three variables lead to even less confidence in the influence of those variables in program evaluation behaviour. The ANOVA described in Section 6.7.4 is likely a better test for indicating the effect of the variables on evaluation behaviour.

6.7.4 Tests of differences between levels on Bennett's Hierarchy

One-way Analysis of Variance with Scheffe and Duncan Post Hoc tests were used to test the null hypotheses of no significant differences in means among levels of Bennett's Hierarchy (the variable that measured evaluation behaviour --"ways evaluated most of the time"- survey question 16). For this test, unlike for the regression, the levels of Bennett's hierarchy categorical variable was the independent variable and the characteristics of NREPs (years of experience, age, attitude summated score, and perceived organizational commitment summated score) served as the dependent variables. The variable, 'ways evaluated most of the time' was condensed from six to four groups (asking participants for reaction at end of event, measuring changes in KASAs at end of event, re-contacting participants to access change, and measuring changes in behaviour or long-term conditions) because of the small number of cases in each group. The ANOVA was used to identify significant differences among these groups or levels on the hierarchy (note that they are not the same groups as identified for the cluster analysis). The Scheffe test was selected as the Post Hoc test because it can tolerate high variability in the number of cases per group compared with other Post Hoc tests (e.g. one group had 109 cases while another had 11). However, Duncan's Post Hoc test was used with the "age" variable, because the Scheffe test did not show significant differences for that variable, despite the risk of Type 1 error using Duncan's Post Hoc Test. The ANOVA was conducted using the following continuous dependent variables:

- years of experience;
- age;
- attitude summated score; and
- perceived commitment summated score.

Percent of salary derived from grant sources (also called funding source) was recoded from a continuous variable to a categorical variable because its distribution was non-normal (bimodal), and was not used for the ANOVA. A cross tabulation and chi-square was used to test correlations between the two categorical variables. The summary of the cross tabulation is on Table 6-35.

ANOVA has three assumptions that should be tested in order to ensure confidence in the results, although the procedure is robust to the violations. The three major assumptions and their diagnostic procedures of ANOVA are (Hair et al. 1998):

- Normal distribution of the dependent variables. I observed the histogram overlaid with a normal curve. I conducted a data transformation on 'years of experience' (square root) because it did not have a normal distribution.
- Each of the groups has equal variance. I tested homogeneity by running Levene's test in SPSS.
- The independent variables used in the procedure are not correlated with each other. A correlation matrix showed that there were significant correlations among two independent variables: age and perceived organizational commitment.

As some groups (categories) had fewer than two cases, the ANOVA could not be conducted without collapsing the dependent variable into a smaller number of groups with more cases per group. Thus, I recoded the variable from five into four groups, combining behaviour changes and long-term changes into one category. The remaining groups were:

1) asking participants for reaction at end of event (called "group 1");
2) measuring changes in KASAs at end of event (called "group 2");
3) re-contacting participants to access change (called "group 3") ; and
4) measuring changes in behaviour or long-term conditions (called "group 4").

The ANOVA results are summarized in Table 6-45 (alpha = 0.05). The results suggest that attitude, years of experience, and age influence program evaluation behaviour, but perceived organizational commitment to evaluation does not, because the *p value* of perceived organizational commitment is greater than .05.

Table 6-45 Statistics on differences between groups (levels on Bennett's hierarchy)

		Sum of Squares	**df**	**Mean Square**	***F Statistic***	***p value***
Attitude toward program evaluation						
	Between groups	4.052	3	1.351	4.072	.008
	Within Groups	67.671	204	.332	-	-
Years of Experience *(transformed - square root)*						
	Between groups	891.49	3	297.16	3.709	.013
	Within Groups	16,185.21	202	80.13	-	-
Perceived organizational commitment						
	Between groups	3.40	3	1.14	2.44	.066
	Within Groups	95.12	204	.466	-	-
Age						
	Between groups	872.13	3	290.71	3.291	.022
	Within Groups	17,315.74	196	88.34	-	-

1) Years of experience

Figure 6-16 shows which groups or levels of program evaluation have significant differences with regards to years of experience. There are no significant differences between groups 2, 3, or 4 but there are differences between group 1 and group 4. In other words, there is a significant difference in years of experience among those who reported that they ask participants for reactions at the end of events most of the time, and those who measure behaviours or long-term conditions. This indicates that there is an effect of years of experience in Extension on level of program evaluation conducted most of the time. Those with more experience in Extension conduct evaluation at higher levels most of the time.

	Measure changes in KASA at end of event (group 2)	Ask participant for reaction at end of event (group 1)	Re-contacted participants to assess changes (group 3)	Measure changes in behaviour or long-term conditions (group 4)
Means (years)	11.36	12.33	16.15	18.65

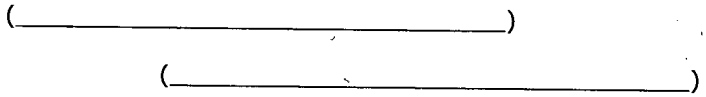

Figure 6-16 Means for years of experience

Levene's homogeneity statistic is .864 with $p = .46$ and three degrees of freedom which shows that the variances of the four groups are equal, satisfying an assumption of ANOVA. The *F statistic* is 3.71 with .013 significance, indicating that there are significant differences at the .05 significance level among the four groups with respect to years of experience.

The Scheffe Post Hoc test shows that there are significant differences between group 1 and group 4 (mean difference = 7.28 with .032 significance) with respect to years of experience.

While the regression did not show a strong linear relationship between evaluation behaviour and years of experience, the ANOVA shows that years of experience does have an effect on program evaluation behaviour. Those who measure higher levels of evaluation have a higher average number of years of experience. The cluster analysis showed a slightly higher number of years of experience for "keeners" (those who do higher level evaluations than for "middle roaders" or "foot-draggers"), and this test shows that the difference is indeed significant.

2) Attitude toward program evaluation (summated score, where 1 = strongly agree and 5= strongly disagree with positive statements about evaluation)

	Re-contacted participants to assess changes (group 3)	Measure changes in KASA at end of event (group 2)	Ask participant for reaction at end of event (group 1)	Measure changes in behaviour or long-term conditions (group 4)
Means (attitude score)	2.20	2.37	2.61	2.67

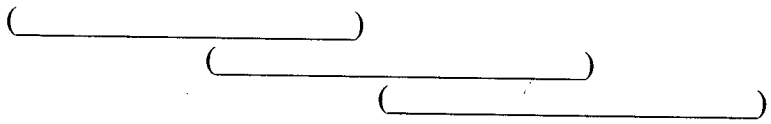

Figure 6-17 Means for attitude summated score

Levene's homogeneity statistic is 1.56 with $p = .20$ and three degrees of freedom. This shows that the variances of the four groups are equal, satisfying an important assumption of ANOVA. The *F statistic* is 4.07 with $p = .008$, revealing significant differences between the groups in the model. In other words, attitude towards program evaluation has an effect on behaviour.

Scheffe's Post Hoc test indicates that there is a significant difference between groups 3 and 1 and groups 2 and 4. There are no significant differences between groups 3 and 2, group 2 and 1, and group 1 and 4.

3) Perceived organizational commitment (summated score, where 1 = strongly agree and 5= strongly disagree with positive statements about organizational commitment)

	Ask participant for reaction at end of event (group 1)	Measure changes in KASA at end of event (group 2)	Re-contacted participants to assess changes (group 3)	Measure changes in behaviour or long-term conditions (group 4)
Means	2.58	2.33	2.62	2.67

Figure 6-18 Means for perceived organizational commitment

Levene's test of homogeneity statistic is 0.67 with p= 0.57 and three degrees of freedom, showing that the assumption of equal variance is met. The *F statistic* is 2.44 with p= .066 showing that there are no significant differences between groups. In other words, perceived organizational commitment does not have an effect on evaluation behaviour. The cluster analysis showed that those NREPs conducting higher level evaluations tended to have

slightly higher average scores for perceived organizational commitment, but this test shows that that was not significant.

4) Age

	Ask participant for reaction at end of event (group 1)	Measure changes in KASA at end of event (group 2)	Re-contacted participants to assess changes (group 3)	Measure changes in behaviour or long-term conditions (group 4)
Means	45	44	51	50

(_____)
(_____)

Figure 6-19 Means for age

Levene's homogeneity of variance statistic is 1.05 with 0.370 significance and three degrees of freedom so assumptions of homogeneity of variance are met. *F statistic* is 3.29 with $p = .022$ which shows that there are significant differences between groups with regards to age. The Duncan Post Hoc test was used to identify which groups were different because the Scheffe and Tukey tests were not sensitive enough to show differences.

The Post Hoc test showed that there were no significant differences between groups 1 and 2, but together they were different than groups 3 and 4. The mean age of those conducting higher levels of evaluation is higher than the mean age of those conducting lower-level evaluations. In other words, older people are more likely to conduct higher levels of program evaluation than younger people.

6.7.4.1 Conclusions about influence of attitude, years of experience, perceived organizational commitment, and age

The ANOVAs showed that some factors have an influence on a NREP's level on Bennett's Hierarchy at which they typically evaluate programs, but some do not. NREPs who have high positive attitude scores conduct higher levels of evaluations. The notion that NREPs who perceive that their organizations are committed to program evaluation conduct higher levels of evaluations is not substantiated by the ANOVA. Older, more experienced NREPs tend to conduct higher levels of evaluation. From this series of tests, it appears that evaluation behaviour is independent from perceptions regarding organizational commitment to evaluation; behaviour appears to be more a function of a person's age, and years of experience, and attitude towards evaluation than other factors.

6.7.5 Differences among tenured vs. non-tenured regarding program evaluation behaviour (tenure status)

I ran an independent samples t-test to determine if there was a significant difference between tenured and non-tenured NREPs with regard to program evaluation behaviour (using summated survey question 15). The t-test showed that there is no significant difference, as shown in the tables below ($p = .954$).

Table 6-46 T-test for tenure status by evaluation behaviour

	N	Mean	Std. Deviation	t-value	Sig.
Have achieved tenure	59	3.733	1.2412	.058	.954
Have not yet achieved tenure	22	3.716	1.0273		

Mean difference = .017
Levene's test- equal variances assumed F= 2.342, p= .130

6.7.6 Differences among those who believe their performance is assessed on basis of program evaluation and those who believe it is not

I ran an independent samples t-test using the summated survey question 15 as the grouping variable to determine if there was significant difference between those who believe their performance is assessed on basis of program evaluation and those who believe it is not. The t-test showed that there was a significant difference, as shown in the tables below. Those who believed their performance was assessed on the basis of program evaluation conducted evaluation at a higher level than those who did not believe their performance was assessed on the basis of program evaluation ($p = .001$).

Table 6-47 Mean differences between those who believe their performance is assessed on basis of program evaluation and those who do not believe it is not-

	N	Mean	Std. Deviation	t-value	Sig.
Believes performance is assessed on whether or not conducts evaluation	89	4.065	1.1552	3.397	.001
Believes performance is not assessed on whether or not conducts evaluation	95	3.487	1.1511		

Mean difference = .578
Levene's test- equal variances assumed F= .268, p= .606

6.7.7 Difference between those who have access to evaluation specialists and those who do not

I ran an independent t-test to determine if there was a significant difference between those who had access to evaluation specialists than those who did not. This test shows that those who had access to evaluation specialists, in fact, performed at a higher level of program evaluation than those who did not ($p = .000$)

Table 6-48 Those who have access to evaluation specialists and those who do not.

	N	Mean	Std. Deviation	t-value	Sig.
Have access to evaluation specialists	66	4.220	1.2584	3.753	.000
Does not have access to evaluation specialists	146	3.574	1.1139		

Mean difference = .646
Levene's test- equal variances assumed F= 1.829, p= .178

6.8 Summary of quantitative results

The table below summarizes the statistical procedures used and the results of each variable as it influences program evaluation behaviour. A "Yes" means that the variable does have at least some influence on program evaluation behaviour, a "No" means that it has no significant influence. A blank cell means the test was not applicable nor conducted for that variable.

Table 6-49 Variables and their influence on program evaluation behaviour.

"Yes" indicates that the variable influences program evaluation behaviour *"No" indicates that the variable does not influence program evaluation behaviour*	**ANOVA and Independent Samples T-tests** (where indicated)	**Multiple linear regression**	**Cross tabulation with chi-square statistic**
VARIABLES RELATED TO HYPOTHESES			
1) Attitude about evaluation	**Yes** (ANOVA)	**Yes-** but weak	N/A

Table 6-49, cont. Variables and their influence on program evaluation behaviour.

"Yes" indicates that the variable influences program evaluation behaviour *"No" indicates that the variable does not influence program evaluation behaviour*	**ANOVA and Independent Samples T-tests** (where indicated)	**Multiple linear regression**	**Cross tabulation with chi-square statistic**
2) Perceived organizational commitment	No difference (ANOVA)	**Yes-** but weak and was correlated with other independent variables	N/A
3) Age	**Yes-** older= higher eval (ANOVA)	**Yes-** weak but strongest variable in model	N/A
4) Years of Experience	**Yes-** more experience = higher eval (ANOVA)	Dropped out of model	N/A
5) Source of funding (categorical)	N/A	N/A	No
6) Position classification (tenure/non-tenure)- dummy variable	N/A	**Yes-** but weak and was correlated with other independent variables	No
7) Tenure status- (dummy variable in regression)	No- (T-test)	Dropped out of model	No
8) Believes performance is assessed on basis of program evaluation (dummy variable in regression)	**YES** (T-test)	**Yes-** but weak and was correlated with other independent variables	No
OTHER FACTORS TESTED			
Have access to evaluation specialist	**Yes-** (T-test)	N/A	N/A
Job classification (county level/state level) (dummy variable in regression)	N/A	Dropped out of model (dummy variable)	**Yes**
Confidence level	N/A	N/A	**Yes**

Although the purpose of the cluster analysis was to classify, not to predict, the cluster analysis supported the finding of confidence level, access to an evaluation specialist, and belief that performance is assessed on basis of program evaluation as factors influencing program evaluation behaviour.

I expected that the quantitative data would provide insight into what makes NREPs more likely to conduct higher levels of program evaluation and be useful to administrators who are looking for ways to induce NREPs to increase the level of program evaluations. It appears from this analysis that the factors influencing evaluation behaviour are associated with both intrinsic (personal) attributes (age, years of experience, and attitude) and extrinsic factors (belief that performance appraisal is assessed on basis of program evaluation and access to evaluation specialists). While confidence level is intrinsic, it can be influenced through interventions such as training.

Table 6-50 relates back to the hypotheses listed in Section 4.0, with the addition of a column indicating whether or not the hypothesis was substantiated. It shows that the hypotheses related to attitude, perceived organizational commitment, age, years of experience in the profession, job classification, and belief that performance appraisal is based on program evaluation behaviour were substantiated, while the hypotheses related to funding source and tenure status were not substantiated. One "Δ" means that a relationship existed, albeit weak, and three "$\Delta\Delta\Delta$" means that the finding was very substantial and provides more confidence in the results.

Table 6-50 Hypotheses and findings

(one Δ means hypothesis is substantiated but conclusion is weak)
(three ΔΔΔ means hypothesis is fully substantiated)

Hypothesis	**Finding**	Conclusion
There is a positive linear relationship between "**attitude toward program evaluation**" and "level of evaluation conducted by NREPs most of the time in the past 12 months." The higher the attitude score of NREPs toward program evaluation, the higher the level of evaluation that has been conducted by NREPs most of the time in the past 12 months.	The ANOVA showed that attitude is a factor in program evaluation behaviour. NREPs with positive attitude toward evaluation conduct higher levels of evaluation. Multiple linear regression shows weak but significant linear correlation between attitude and program evaluation behaviour.	ΔΔΔ
There is a positive linear relationship between "**perceived organizational commitment**" and "level of evaluation conducted by NREPs most of the time in the past 12 months." The higher the perceived commitment of the organization to evaluation, the higher the level of evaluation that has been conducted by NREPs most of the time in the past 12 months.	The ANOVA showed that perceived organizational commitment was not a factor in evaluation behaviour. Multiple linear regression showed weak but significant linear correlation between level of perceived organizational commitment and program evaluation behaviour.	Δ
The mean level of evaluation conducted by NREPs most of the time in the past 12 months differs by **age** (older NREPs conduct higher levels of evaluation than younger NREPs).	The ANOVA showed that age is a factor in program evaluation behaviour. Older NREPs conduct higher levels of program evaluation than younger NREPs. Multiple linear regression showed weak but significant linear correlation between age and program evaluation behaviour.	ΔΔΔ
The mean level of evaluation conducted by NREPs most of the time in the past 12 months differs by **years of experience** in the Extension profession. (NREPs with more years of experience conduct higher levels of evaluation than less experienced NREPs).	The ANOVA shows that NREPs with higher years of experience conduct higher levels of program evaluation.	ΔΔΔ

Table 6-50, cont. Hypotheses and findings

(one Δ means hypothesis is substantiated but conclusion is weak)
(three ΔΔΔ means hypothesis is fully substantiated)

Hypothesis	Finding	Conclusion
The mean level of evaluation conducted by NREPs most of the time in the past 12 months differs by **funding source**. (NREPs whose positions have a higher percentage of grant/soft funds conduct higher levels of evaluation).	The ANOVA, multiple linear regression and cross tabulation/chi-square showed that percent of soft funding for salary does not appear to influence program evaluation behaviour. The cluster analysis also showed that funding source is not different between keeners, middle roaders and foot-draggers. The chi-squared statistic between funding source as a categorical variable and survey Question 16 does not indicate significant association.	NO
The mean level of evaluation conducted by NREPs most of the time in the past 12 months differs by **position classification**. (NREPs whose positions are considered "tenure track" have conducted higher-level evaluations in the past 12 months than those whose positions are not "tenure track").	The multiple linear regression showed that position classification has a weak but a significant influence on program evaluation behaviour. The chi-square statistic also shows a significant relationship using survey Question 16 and position classification (using a dummy variable)	Δ
The mean level of evaluation conducted by NREPs in the past 12 months differs by **tenure status**. NREPs whose positions are "tenure track" but who have not yet achieved tenure status have conducted higher-level evaluations in the past 12 months than those whose positions are tenure-track but who have already achieved tenure status.	No, the independent samples t-test to test differences in means between tenured and non-tenured respondents showed no significant difference.	NO
The mean level of evaluation conducted by NREPs in the past 12 months differs by **whether they believe their personnel performance appraisal is based on the basis of program evaluation behaviour**. (NREPs who believe their personnel performance is based on the level of program evaluation conducted conduct higher-level evaluations than those who don't believe their personnel performance is based on program evaluation conducted).	Yes, an independent sample t-test using summated survey Q15 showed that those who believe their personnel appraisal is based on whether or not they conduct program evaluation did conduct higher levels of evaluation most of the time in the past 12 months. The regression shows that there is a linear relationship between belief that personnel performance appraisal is based on whether or not they conduct program evaluation and program evaluation behaviour, albeit a weak relationship.	ΔΔΔ

In addition to factors discussed in my hypotheses, three other factors were found to influence evaluation behaviour as show in table 6-51.

Table 6-51 Other factors found to influence evaluation behaviour.

(one Δ means hypothesis is substantiated but conclusion is weak)

(three ΔΔΔ means hypothesis is fully substantiated)

Variable	Explanation	Conclusion
Have access to evaluation specialist	T-test showed significant differences in evaluation behaviour between respondents with access to an evaluation specialist and those without access to an evaluation specialist.	Δ
Job classification (county level/state level)	Cross tabulation and chi-square statistic showed significant influence of job classification on evaluation behaviour. A higher percentage of state level NREPs conduct higher levels of evaluation but about the same percentage re-contacts participants after events.	Δ
Confidence level	Cross-tabulation and chi-square statistic showed significant influence of confidence level on evaluation behaviour. Respondents who felt more confident conducted higher levels of evaluation.	Δ

7.0 QUALITATIVE RESULTS

Three survey questions asked respondents to provide qualitative information on barriers to program evaluation (survey question 13, Appendix C), approaches to evaluation beyond the end-of-event questionnaire (survey question 4a, Appendix D), and general comments and recommendations for improving the practice of evaluation within their organizations (survey question 24, found in Appendix E).

7.1 Barriers to evaluation

Many barriers mentioned by respondents related to each other (such as lack of skills related to lack of organizational system for building evaluation skills) but could be categorized into the following:

-lack of skills. Eighteen respondents described lack of skills as a key barrier. Seven respondents who mentioned lack of skills also pointed to other barriers that kept them from increasing skills, such as lack of time for learning, lack of leadership for professional development activities, and infrequent training opportunities.

> *"Most of us working in county programs are Master's level. We don't have a deep well of scholarly evaluation methods from which to draw."*

-time prioritization issues. Twenty-eight respondents spoke about lack of time and other priorities that kept them from conducting in-depth or meaningful evaluations. Most referred to too many expectations and being "spread so thin."

> *"Once a project ends, it's on to another. I am personally motivated to do continuous improvement, but the scope and complexity of my day-to-day responsibilities limits my ability to devote much time to evaluation."*

-lack of funding. Fourteen respondents mentioned lack of funding as a barrier. One respondent referred to being funded by soft money that provides no budgets for evaluation. Several referred to staff cutbacks that limited their time for evaluation.

> *"Good evaluations can be expensive. Our resources are constantly being cut but we are still expected to evaluate. When staff and budgets get cut, evaluation sometimes gets dropped."*

-methodological difficulties. A variety of barriers related to evaluation methods was mentioned, including client survey fatique (too many surveys), inability to interpret data

collected, low survey response rates, social desirability bias ('clients telling you what you want to hear'), tedious human subjects review processes, difficulty assessing behavioural changes and difficulty getting contact information for clients.

"In many cases the program we want to evaluate is not conducive to follow up (difficulty accessing people)."

-lack of organizational system for evaluation. Two respondents mentioned lack of an official evaluation system in their organizations.

"The organization is aware and wanting us to do more significant evaluations (more explicit impacts) but not yet giving us all the means to make the transition."

-lack of rewards for doing evaluation/lack of administrative incentives. Thirty respondents described a host of barriers related to lack of rewards, including training provided without follow up or support, belief that evaluation is viewed as "non-productive" by administrators, lack of guidelines for evaluation, inconsistency among organizational leaders regarding the expressed value of evaluation, belief that the tenure system works against evaluation because there is little incentive to conduct research outside one's area of expertise.

"Evaluation is easy to leave behind, since the reward for doing so is not so great. I have seen very little evidence that this is a real priority for my organization in that time and money are not earmarked for evaluation. It's expected and required but not well supported."

-scepticism about the value of formal evaluation. Twenty-one respondents were sceptical about the value of conducting evaluation and offered a wide range of explanations.

"Evaluations of my presentations are sufficient."

"So many of the evaluations I have seen others do seem valueless. "

"Word of mouth is a better gauge of effectiveness."

"Disinterest is my only barrier."

"The idea that we are responsible for our client's actions is politically driven and in some ways unrealistic."

"I suspect that some folks are too nice to critically evaluate a program."

7.2 Approaches to evaluation beyond the end-of-event questionnaire

Respondents provided examples of approaches they used other than questionnaires distributed to participants at the end of Extension events. These included methods such as:

-needs assessments

-methods of measuring KASA (Knowledge, attitudes, skills, and aspirations) such as pre and post tests, surveys to assess perceptions of increased forest stewardship skills

-methods of measuring practice/behavioural changes, such as follow-up interviews and surveys, observation of participants who had hired a forester, had a stewardship plan, or prepared or harvested timber, and phone calls to participants one year after attending workshop asking about behavioural change

-higher level impacts such as measuring cost reduction, improved utilization, and job creation

-informal evaluation such as advisory council, summaries of conversations, including audience members in planning, informal discussions with program participants

-other methods, such as focus groups of past participants, web-based surveys.

7.3 Comments and recommendations from respondents

Sixteen respondents provided comments on the value of evaluation to them. Most responses reflected an understanding of the importance of evaluation but some were qualified with comments such as "*....I am not a big fan of evaluation but I understand its importance*" and "*I don't like doing evaluation, but I value the outcomes of it,"* and "*evaluation tools have a value if there is no threat of them being misued for personal advancement."* Respondents who provided comments on evaluation spoke more about its value to them for program planning than for accountability.

Recommendations for improving the practice of evaluation in their organization covered a range of topics, including:

-administrative incentives/rewards/prioritization for evaluation. Twenty respondents commented on the importance of incentives and several suggested that the incentives be positive reinforcement such as recognition and monetary rewards rather than negative inducements. Two respondents stressed the importance of explaining how and why evaluations are important, rather than just requiring them. Another respondent suggested providing success stories to illustrate how evaluation strengthens programs to shift evaluation

from a "must do" activity to a "want to do activity." Several respondents suggested tying evaluation to performance and promotion and tenure assessments.

> *"Emphasize the importance of evaluations to be sure that we are satisfying stakeholder needs."*

-standardization/establish evaluation system. Forty-two respondents provided comments related to the need to establish systems within their organizations for evaluation such as policies that include guidelines regarding the number and type of evaluation expected each year, *"standard, easy-to-use, convenient"* evaluation tools (such as survey templates) and processes.

> *"For every program there should be an evaluation strategy on the web page."*

-access to evaluation expertise and resources. Nineteen respondents suggested a need for evaluation specialists to assist and coach them. Four respondents suggested providing reference materials and websites.

> *"In an idea situation, there would be an evaluation office – a group of evaluation specialists that would work with Extension faculty to design, administer, analyse, and summarize evaluations."*

-mandatory and optional training. Thirty-one respondents suggested training, but only two suggested that the training should be mandatory. Suggested topics included statistics, databases, and evaluation methods. One person stressed a preference for self-paced, distance-delivered, user-friendly training and another stressed that it should be practical with tips and templates.

> *"It needs to be more than a two-hour lecture on how to do a particular method while another said, " it should be accomplishable in a two hour training period."*

-funding dedicated to evaluation. Six respondents commented on the need for funding.

> *"With support funding, which would purchase time to do evaluation, we would do more."*

-include evaluation results in planning. Two respondents spoke of a desire to see the results of their evaluations used at state levels.

> *"Show examples of how a program evaluation substantively improved a program or increased support for the program."*

-Other comments

-more sharing of tools among programs
-develop a newsletter on evaluation and feature successful program evaluations in each issue

7.4 Summary of qualitative responses

Respondents offered ideas on barriers, descriptions of evaluation approaches used by NREPs beyond end-of-event questionnaires, as well as comments and recommendations to Extension administrators in the qualitative portion of the survey. They mentioned a wide range of barriers that related to both intrinsic and extrinsic factors that generally supported the quantitative findings. Recommendations also focused on both intrinsic (mainly training to increase skills) and extrinsic factors (standardization and incentives, dedicated funding. There did not appear to be inconsistencies between the quantitative and the qualitative findings relating to factors that influence evaluation behaviour.

8.0 DISCUSSION

The hypotheses used in this study are derived from my experiences and observations as an Extension professional regarding my evaluation behaviours and those of my Extension colleagues. From experience, I assumed that NREPs who were tenure-track but had not yet achieved tenure would conduct higher levels of evaluation compared with their tenured colleagues who no longer needed to conduct research to achieve tenure status (I expected that higher-level evaluation was one way to satisfy the research requirement). I also expected that administrators in states in which NREP positions were non-tenure track would have more difficulty motivating their personnel to conduct evaluation because of the absence of the tenure goal. However, the data surprised me by showing no relationship between tenure status and evaluation behaviour, and only a questionable link between position classification (tenure track, non-tenure track) and evaluation behaviour.

It is possible that tenure-track NREPs who are not yet tenured are more motivated to conduct *research* than tenured NREPs, but they may not consider evaluation as a way of fulfilling research requirements, or they find easier or more desirable types of research than evaluation to fill their dossier. One respondent comment supports this possibility:

> *"The tenure system works against evaluation. There is little incentive to complete essentially what is research that is outside my expertise for which I can be rewarded."*

It is also possible that non-tenure track NREPs are motivated to conduct evaluation to the same degree as tenure-track NREPs for different reasons such as grant requirements or because they simply want to know the impacts of their work for their own satisfaction.

Overall, the data revealed that only some of my hypotheses were fully substantiated. Only age, years of experience, attitude, and belief that performance appraisal is based on evaluation behaviour were strong predictors of evaluation behaviours, while, perceived organizational commitment and position classification were weaker predictors of evaluation behaviour. Tenure status and funding source did not appear to have any influence on evaluation behaviour. The qualitative data showed that a number of respondents discussed the importance of performance appraisals and administrative expectations in their behaviour and approach to program evaluation.

I expected perceived organizational commitment to score higher as a predictor of evaluation behaviour than it did. This survey question was an index question, and it's possible that some of the sub-questions selected were not precise indicators of perceived organizational commitment. There was not an index available in the literature for measuring perceived organizational commitment toward evaluation so I was forced to develop a empirically based index that may have skewed the results. However, the Cronbach alpha

reliability test would have eliminated sub-questions that would have thrown off the results in the index question, so I am quite confident in the findings regarding that variable.

In addition to the variables in the original hypotheses, I found that other variables had an influence on evaluation behaviour. The data showed that "confidence level in conducting evaluation" and "access to an evaluation specialist" were related to higher levels of evaluation. The qualitative data showed that many respondents listed lack of skills as a key barrier to program evaluation behaviour. More than 10 recommended access to evaluation specialists or statisticians for improving the practice of evaluation in their organizations. These are notable findings because they point to two tangible interventions that Extension administrators can implement to enhance evaluation behaviours: training and mentoring to increase confidence, and hiring evaluation specialists to assist NREPs.

I used ANOVA, multiple linear regression, and cross tabulations/chi-square to examine relationships between personal and job characteristics and evaluation behaviour; there were some inconsistencies in the results of these tests with the same variables. For example, the ANOVA showed there were no differences in perceived organizational commitment among those conducting various levels of evaluation most of the time, while the regression showed a weak but positive linear correlation between variables that included perceived organizational commitment (summated score) and evaluation behaviour (r^2 = .223).

The large number of qualitative responses (over 140 individual comments) on barriers and recommendations shows that NREPS have many thoughts and ideas regarding "what's wrong" with the current situations regarding evaluation and how to improve the practice. The largest number of recommendations were in the categories "standardization/establish evaluation system" and "administrative incentives/rewards/prioritization for evaluation." Many respondents felt that verbal encouragement or requirements for evaluation were not backed up with adequate support systems, guidelines, or administrative procedures. The strongest statements seemed to be related to time prioritization—many respondents spoke of too many other program priorities that took precedence over evaluation.

The ANOVA results are more robust than the findings of the regression because of the weak r^2 value of the regression and the correlations among independent variables, but the inconsistency is still a curious finding. There are several possible explanations for this inconsistency. I used two different survey questions (survey questions 15 and 16) as indicators of evaluation behaviour which may have accounted for the difference between the ANOVA and regression conclusions. Or, as stated above, perhaps not all sub-questions used for measuring perceived organizational commitment were good indicators of organizational commitment. A third plausible explanation is the complexity of human perceptions that leads people to report differently than they behave.

Most of the previous studies on Extension evaluation attitudes and behaviours (including Douglah et al. 2003, Jha 2001, Radhakrishna and Martin 1999, Summerhill and Taylor 1986) focused on one geographic region and did not look at variables that existed

across states or within one discipline. Still, the findings of this study are consistent with several previous studies on some points. There is consistency regarding perceived barriers (lack of time, skills, and incentive) and the wide range of evaluation attitudes and behaviours among respondents between this study and earlier studies.

The findings regarding age and years of experience as predictors of evaluation behaviour in this study are consistent with Summerhill and Taylor's 1986 study of Extension agents in Florida. Simply, the longer one is in the profession, the more likely one is to conduct evaluation. The fact that confidence and skill levels appeared as an important factor is also not a surprise; other studies and papers have also identified this factor (Summerhill and Taylor 1986, Douglah et. al 2003).

To my knowledge, however, this is the first study that establishes the importance of evaluation specialists and personnel performance appraisal criteria as significant factors related to evaluation behaviour.

8.1 Motivational theories and the variables influencing evaluation behaviour

Motivational theories discussed in Section 2.9.1 offer some explanations for the findings. Each variable found to influence evaluation behaviour in this study can be linked to a motivational theory, as shown in Table 8-1.

Table 8-1 Variables influencing evaluation motivation and related motivation theories

Variable	Motivation theory
Age/years of experience	Life course theory
Position classification	Behaviourist theory
Performance appraisal based on evaluation	Behaviourist theory
Access to evaluation specialist	Behaviourist theory, Self-efficacy theory
Confidence level	Self-efficacy theory
Attitude toward evaluation	Personal investment theory and socialization theory

Life course theory would predict that an individual's motivation for evaluation would be influenced by their age and stage of career. Older people, for example, might be motivated by a desire to increase the meaning of their life's work by trying to uncover the "so what?" of their efforts. More experienced NREPs who might be more settled in their careers might be more motivated to seek an understanding of the impacts of their life's work than less experienced NREPs.

Behaviourist theory would predict how an individual could be moved to conduct evaluation if they perceive there are rewards associated with positive performance appraisals. Individuals may conduct evaluation as long as it appears that their performance is assessed on the basis of their evaluation behaviour but would decrease or eliminate their evaluation behaviour if/when they move to a job in another state where this did not exist.

Self-efficacy theory explains how motivation to conduct a task is directly related to one's perception of their ability to accomplish it. According to this theory, NREPs who feel confident about conducting evaluation are also more likely to be motivated to conduct it.

Personal investment theory explains how positive attitude toward evaluation could motivate an individual to conduct evaluation because conducting evaluation (and, presumably, realizing the results of evaluation) provides meaning or value to them personally.

Socialization theory would predict that an individual's university training influences their attitude towards evaluation. Those who received their training at a higher-education institution that emphasized and valued evaluation and that teaches and expect students to have competencies in evaluation are more likely to produce students who value and emphasize evaluation in their work life.

Figure 8-1 shows the factors revealed in this study and their corresponding theory from the literature. The dark shade shows factors that are stronger indicators.

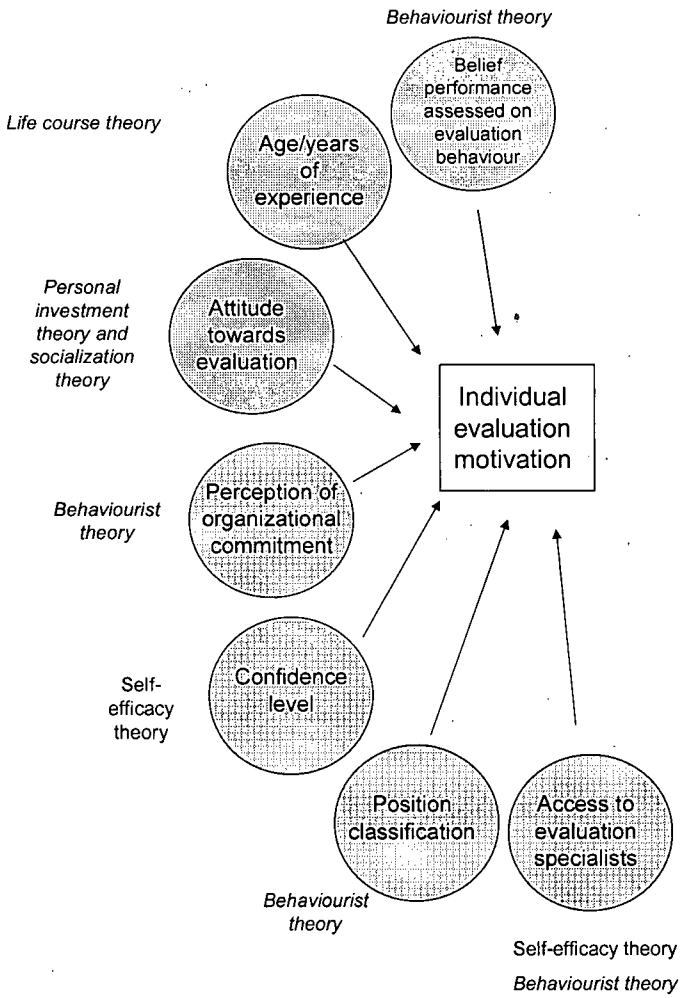

Figure 8-1 Factors affecting individual evaluation behaviour and their theoretical foundations

8.2 Individual motivation and organizational Evaluation Capacity Building

If the task of this thesis was strictly to establish the factors that influence the behaviour of Extension professionals in one organization regarding evaluation, then the discussion would finish here. I would end by suggesting that if Extension organizations could hire only experienced people who feel confident and have positive attitudes about evaluation, if those people were given access to evaluation specialists, and it was ensured that their performance was assessed on the basis of their evaluation behaviour -- then their evaluation worries would be over. However, the story of how well an organization ultimately performs at evaluation is much more complex than only focusing on motivation of employees at one moment in time; the findings need to be tied to a larger discussion on the dynamics of organizational evaluation capacity building and the factors that make evaluation viable within an organization over time. How do organizational functions lead individuals to conduct evaluation, and how does individual behaviour link to organizational functioning? Much of the literature on evaluation capacity building (ECB) has been published since the earlier Extension studies were completed, so it is important to introduce the dynamic interaction between individual and organization that has been addressed very little in the Extension evaluation literature to date.

The recent literature on ECB reveals that there are many interactive factors related not only to individual motivation, but also to organizational systems and processes that influence the capacity of an organization to effectively evaluate its work. Individuals themselves are complex, and they operate in complex social situations in complex environments. The ECB literature would instruct us to look at not only what makes NREPs tick regarding evaluation, but what makes an organization tick regarding evaluation, with the individual as one – but only one –component.

The original proposal for this research was aimed at helping administrators create effective incentives and disincentives and recruitment standards to induce more and a higher level evaluation by employees. My thinking matured as I read the literature on ECB and its related concepts (mainstreaming, enculturation etc). I had assumed that improved program evaluation was a matter of providing incentives and disincentives such as those described in Table 2-5 and that NREPs would magically respond. However, non-cognitive iapproaches to work motivation such as incentives and disincentives are hotly debated in the literature. Many theorists now agree that rewards and disincentives may lead to movement, but not to motivation, and are thus unlikely to lead to sustained behaviours. Remove the external force and the behaviour stops, or so the theory goes. Recent academic thinking about work motivation is well articulated by Kohn (1999) when she said, "If our goal is excellence, no artificial incentive can ever match the power of intrinsic motivation. Rewards, like punishment, may actually undermine the intrinsic motivation that results in optimal performance."

At the same time, theorists also agree that no single motivation theory can alone account for work behaviours. The results of this study show that the strongest predictors of NREPs evaluation behaviour include the interaction of personal and environmental factors, as illustrated in Figure 8-2.

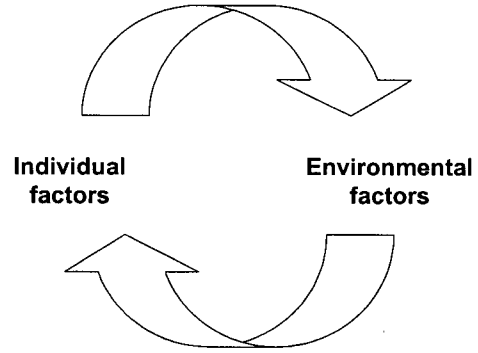

Figure 8-2 Interactive personal and environmental factors that affect program evaluation behaviour

If many theorists agree that non-cognitive approaches are ineffective at creating real motivation, why do most organizations still use these methods? Non-cognitive approaches are easier to implement and measure. It is far more difficult to create an environment where intrinsic motivation for evaluation has the greatest opportunity to surface and blossom, because it involves a comprehensive examination of the organization's structures and cultures and an evolution to a "learning organization" as defined by Senge (1990) and Torres and Preskill (2001) . The literature on evaluation capacity building indicates, however, that this evolution is the key to the sustained use of evaluation throughout the organization. The discussion on how this evolution takes place in organizations is still emerging.

A little more than half of all respondents in this study said that at least 50% of their motivation to conduct evaluation was due to external factors. The greatest number of respondents said that it was 50-50 but the next most frequently selected percentage was 80% external motivation factors. The large amount of qualitative data that discuss administrative rewards also show that a lot of people still rely on incentives and disincentives. If CSREES and state Extension administrators desire for a majority of NREPS to say that all of their motivation for evaluation is internally derived (rather than the current 4.1% of respondents in this survey who said so), then they will have to create the environment for this to happen.

The literature offers an alternative approach to reliance on incentives and disincentives related to the work environment. Cognitive motivation theories assert that the process of individual cognitive goal setting is a precursor to motivation. The most frequently selected source of motivation for evaluation cited by respondents was "seeing evaluation results used in decision making." If NREPs receive feedback that their evaluation results

influence higher-level decision making and make an important contribution to the organization beyond fulfilling a requirement, then I would expect evaluation behaviour to increase. Without this feedback, the cycle of motivation described by Sims and Lorenzi (1992) (i.e. goal setting to results to feedback to more goal setting) is interrupted. If NREPs can see that there are not just rewards for *doing* evaluation but see how their evaluation affects decisions at higher levels, I would expect greater engagement in evaluation.

To illustrate the connection of individual motivation to evaluation capacity building for NREPs in Extension, I propose a framework that links individual NREP evaluation motivation with other components of evaluation capacity building, synthesized from the literature. As shown in Figure 8-3, individual motivation for evaluation is one of five components of ECB. The other four components I will call individual core competencies, effective incentive systems, the presence of support systems, and internal evaluation processes. Individual core competencies, as discussed in Section 2.5, refers to the knowledge, skills, and abilities that employees require to perform evaluation. Effective incentive systems refer to the incentives and disincentives provided by an organization to encourage evaluation behaviours, such as the performance appraisal system or recognition programs. Support systems refer to tools and resources provided to assist employees successfully conduct evaluation, such as web-based and written evaluation guidebooks, and access to evaluation specialists. Internal evaluation processes refer to administrative infrastructure dedicated to evaluation such as evaluation working groups as well as an evaluation lead or champion in the organization. The fifth component, individual evaluation motivation, as I found in my study, appears to be affected by age, years of experience, attitude towards evaluation, and belief that performance is assessed on evaluation behaviours. Access to evaluation specialists, perceived organizational commitment, confidence level, and position classification may have some influence on levels of evaluation conducted. Tenure status, education level, and funding source do not appear to influence the level of evaluation conducted.

There is overlap among the components. In other words, a given administrative intervention designed to enhance evaluation behaviours among employees (such as performance appraisal systems or dedicated funds for evaluation) may fit into more than one component (such as support systems and incentives). Affecting core competencies can affect confidence levels. The purpose of the framework is to helps reveal various types of interventions to increase overall capacity for conducting evaluation as an organization. The framework illustrates that evaluation capacity building is multi-faceted and requires interventions in more than one area. Administrators need to focus not only on motivating individuals, but evolving into "learning organizations" that allow for intrinsic motivation to flourish.

Although this framework focuses on NREPs, it could be adapted for other organizations to help frame discussions on how to build evaluation capacity.

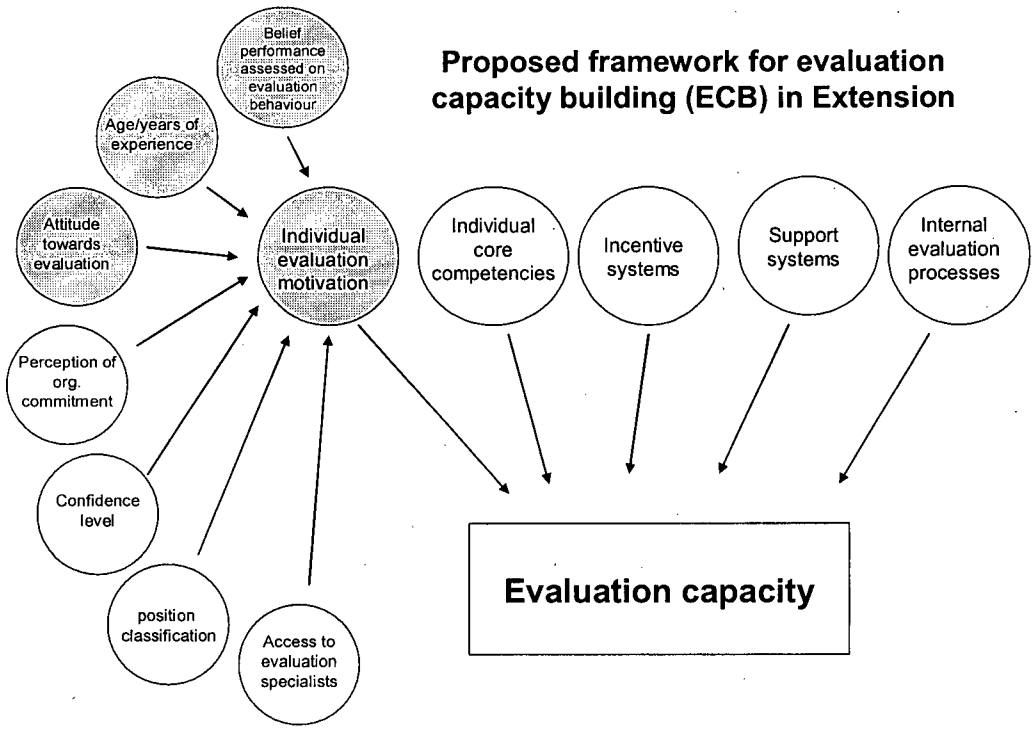

Figure 8-3 Proposed framework for Evaluation Capacity Building (ECB) in Extension

8.3 Implications for ECB in Extension

Known by several names (organizational learning, mainstreaming, enculturation, and ECB), the concept of multi-faceted intervention to enhance evaluation at all levels of an organization needs to continue to be a focus of Extension at the national, state and local levels. Both the quantitative and qualitative data in this study show that Extension organizations need to develop internal evaluation processes, support systems, incentive systems, and enhance individual competencies. The organizations can affect individual motivation through training and mentoring, and by demonstrating commitment to evaluation in ways discussed earlier in this document.

The specific recommendations arising from results of this data are as follows:

1) Hire or maintain evaluation specialist positions to assist NREPs. This study has shown that access to evaluation specialists is linked with program evaluation behaviour. The quantitative data shows that respondents who conducted higher levels

of program evaluation had access to evaluation specialists. Twelve respondents specifically mentioned providing an evaluation specialist as a recommendation for improving evaluation practices in their organizations. Generally speaking, states that can support a dedicated specialist in program evaluation to assist NREPs can expect to have higher levels of program evaluation. While it cannot be determined from this data whether evaluation specialists cause or are only associated with this trend, the presence of an evaluation specialist addresses several of the components of ECB in my framework. Hiring an evaluation specialist communicates organizational commitment for evaluation, can lead to increased confidence levels through training and mentoring, and provides a focal point for improved internal evaluation processes.

2) As a corollary to the above, the literature supports the notion of having evaluation specialists who are committed to and have skills in ECB, not just program evaluation. The skills required for helping build evaluation capacity are different from those for conducting evaluation.

3) The quantitative and qualitative data show that there is merit in the notion of enhancing personnel performance evaluation forms and using them effectively. To ensure the highest level of program evaluation, administrators should include clear criteria relating to program evaluation behaviour and ensure that NREPs are knowledgeable about the criteria. Providing a national standard for performance appraisals regarding evaluation expectations is key to this. There are several excellent examples (Oregon and North Carolina) to follow (see Appendix F). The findings of this study indicate that perceptions regarding performance appraisals that include explicit and clear criteria regarding program evaluation are linked to NREP evaluation behaviour. My analysis also showed that in 16 states, there was disagreement among respondents within the same classification regarding whether or not their performance is appraised on the basis of conducting program evaluation. Several said that they did not know. This indicates that either performance expectations are not clear, individuals in the same job classification work under different performance appraisal systems, or performance appraisal forms were being modified at the time of the survey (a few states indicated that they were planning to update theirs). Extension administrators should be concerned about the fact that some respondents in the same job classification did not agree on whether or not their performance was appraised on the basis of program evaluation. A high percentage (62%) of respondents felt that program evaluation behaviour should be included as a criterion in performance assessment, so there appears to be wide support for doing so.

4) Extension organizations should focus on an organizational cultural mindset and administrative commitment toward evaluation, not just on changing individual

attitudes and behaviours towards evaluation. While this study measured individual motivation, a number of authors asserted that organizational cultures also influence behaviour. Organizations need to adopt evaluation as a core value from the top down, create a common language, and communicate expectations throughout the organization. Douglah et al. (2003) reminded us that "the challenge lies in developing a shared understanding of evaluation that is supported by employees and other stakeholders." Committing appropriate resources for capacity building such as evaluation specialists and resolving ambiguities in performance appraisals are keys to this.

Administrators should be alarmed by comments like one offered by a respondent: "*Taking time for more thorough evaluations would interfere with productivity.*" The assumption of this response is that doing evaluation is not productive and considered an auxiliary project to one's job, not a core part of it. Another respondent said, "*I am in an ongoing rut of preparation to program, to cleanup, and onto the next program. Evaluation seems to be a sidetrack.*" The words "non-productive" to describe evaluation was used by more than one respondent. Lack of time for evaluation was the most frequently mentioned barrier to evaluation, but one respondent conceded, *"Although it would be easy to say "lack of time" is the greatest barrier, if there were a major incentive - e.g., considered in performance appraisal - then I would find the time!"*

Administrators need to demonstrate through various mechanisms that evaluation is an *assumed, expected, and rewarded* part of an NREP's core work functions. These mechanisms could be: job descriptions, recruitment criteria, dedicated funds for evaluation, award programs, merit raises, recognition through featuring evaluation results in newsletters, providing evaluation form templates, standardizing evaluation procedures, providing guidelines about how many and how much and type of program evaluations are expected each year, creating timely, user friendly training and info-sheets, requiring evaluation plans for each project, mandatory and providing training and mentoring programs.

5) Use evaluation results in program decisions. The data revealed that using evaluation results in higher-level program decisions was a key motivator for evaluation behaviour for many respondents. Several respondents recommended demonstrating how their evaluation results are being used in state-level decisions for improving evaluation practice in their organizations. Belief that evaluation is not being conducted just for the sake of conducting the evaluation was important to many respondents. NREPS involved in evaluation should have a local forum for sharing evaluation results with colleagues and decision makers so results can be integrated into planning.

To summarize, based on the findings of this research, Extension administrators could do the following to increase the level of program evaluation being conducted by natural resource Extension professionals: 1) enhance, clarify, and communicate performance appraisal criteria to reflect a priority for program evaluation, 2) hire or maintain evaluation specialists, 3) enhance training and mentoring and resources to increase confidence levels, 4) enhance the standardization and streamline the evaluation process (e.g. create templates, establish guidelines, build into planning process), and 5) develop more forums to seek input from NREPs regarding ways that Extension organizations can nurture intrinsic motivation.

8.4 Recommendations for further research

Several questions still remain about how Extension organizations can incrementally evolve into learning organizations that nurture intrinsic motivation for evaluation. While this study focused on individual motivation, the next step is to look more closely at the other four components of the ECB framework and ask what internal structures, support systems, incentives, and core competencies can lead to improved evaluation behaviour among NREPs.

There is another question that this research did not address: Is motivation to conduct evaluation linked to one's work motivation in general? Is there a correlation between how people feel about their jobs in general and how they feel or behave regarding evaluation? While there are many theories about how to get people motivated at work, I do not know how closely understanding work motivation helps us understand evaluation motivation. It is conceivable that NREPs could be motivated to do other parts of their job, but not evaluation.

Further work is needed to look at how NREPs would characterize organizational commitment to evaluation. The index of sub-questions I used for this study was only based on my best guess.

Most importantly, further research is needed to look at the dynamics of change within organizations over time. As administrators introduce interventions to increase evaluation, for example, what difference does that make? An evaluation of administrative interventions to enhance evaluation behaviour with some case studies would be very informative.

8.5 Summary

This study points to four predictors of program evaluation behaviour for natural resource Extension professionals in the U.S: age, years of experience in Extension, belief that performance appraisal is linked to evaluation behaviour, and attitude about program evaluation. Position classification (i.e tenure-track/non-tenure track), access to evaluation specialists, perceived organizational commitment to evaluation, access to evaluation specialists, and confidence levels in conducting evaluation also may be linked with level of evaluation conducted. The research shows that in general, older, more experienced NREPs with positive attitudes towards evaluation and who believe that their job performance is assessed on the basis of their evaluation practices are linked with higher levels of evaluation.

Tenure-track NREPs with access to evaluation specialists and who perceived their organization is committed to evaluation and who feel confident about conducting evaluation may also be linked with higher levels of evaluation. Tenure status, education level, and funding source for salary do not appear to be linked with level of evaluation conducted.

While Extension administrators have less ability to influence attitude toward evaluation, age, and years of experience, they do have the ability to influence the performance appraisal system, access to evaluation specialists, and confidence levels through training and mentoring programs. They also have the ability to build their organizational evaluation capacity by focusing on internal evaluation processes, support systems, incentive systems, core competencies in addition to focusing on individual factors.

My research will not settle the debate between non-cognitive verses goal theories of motivation. Extension organizations will probably always need incentives to obtain the level of evaluation needed in Extension. However, the literature on motivation, evaluation capacity building, and findings of this research suggest that the interactions between the personal characteristics and the work environment determines NREP evaluation behaviour, and that both need to be addressed in order to enhance evaluation behaviour.

Survey about Program Evaluation -- Natural Resource Extension and Landowner Assistance Professionals

Questions marked with an asterisk (*) are mandatory.

Thank you for participating in this important study about Natural Resource Extension and Landowner Assistance program evaluation in the United States. Our aim is to learn about attitudes and practices of Extension and Service Forestry professionals regarding program evaluation. Please answer the following questions by clicking on the button that most closely reflects your thoughts.

For this survey, "program evaluation" refers to collecting and analyzing data on changes that occur in your target audience as a result of your programs through questionnaires, interviews, observation, e-mail, mail, or telephone surveys. The data can be either quantitative (numbers) or qualitative (anecdotal).

The changes can include increased knowledge and skills, new attitudes, aspirations, or changed behaviors, practices, or social, economic and physical conditions. It refers to evaluations that you conduct individually or as part of a team and ranges from end-of-event questionnaires to longer-term impact studies and needs assessments.

There is a section at the end of the questionnaire for comments.

1 What percent (%) of your current job is devoted to Extension, Service Forestry, or Landowner Assistance activities?

2 Which one of the following most closely describes your current employer? If you work for more than one organization, please select the organization that conducts your personnel performance appraisals.

- Land Grant University – campus-based
- Land Grant University – county/region/district-based
- Non-profit, non-government organization
- Government agency
- Private consulting firm
- Academic institution but not a land grant university
- Other, please specify

3 *Have you conducted any kind of evaluations of your programs during the past year (see definition of evaluation just above Question 1). If you click on "no," then skip to Question 5.

YES NO

4 Have you conducted evaluations of any programs during the past year IN ADDITION TO asking program participants to fill out questionnaires or provide verbal feedback at the end of an event?

YES NO

If yes, please describe

5 Which of the following most closely describes your "gut" reaction to "program evaluation" (measuring the impacts of your programs)?

- I love doing it
- I don't mind doing it
- I would prefer to ignore it
- I absolutely dread it
- I rarely think about it

6 What percent of your motivation to do program evaluation would you say comes from external factors (such as promotion and tenure, grant reporting requirements, and performance appraisals) and what percentage comes from your own personal motivation (such as desire to know if your teaching is effective or desire for professional improvement)? The total should add up to 100 percent.

External (%) _____

Personal (%) _____

7. Please indicate the extent to which you agree or disagree with each of the following statements.

1 Strongly Agree	2 Agree	3 Neither Agree nor Disagree	4 Disagree	5 Strongly Disagree
Program evaluation is deeply rooted in the administrative values of my organization.	1	2	3	4	5
I am not expected to conduct evaluations of my programs.	1	2	3	4	5
My organization positively rewards Extension/Landowner Assistance professionals for conducting program evaluations.	1	2	3	4	5
My administrators are not committed to building evaluation capacity within the organization.	1	2	3	4	5
Program evaluation is a normal part of my organization's culture and practice.	1	2	3	4	5
There are adequate resources within my organization to assist me with program evaluation when I need it.	1	2	3	4	5

8 Which of the following evaluation resources are available to you as an Extension/Landowner Assistance professional in your state? (select all that apply)

- [ ] In-service training in program evaluation
- [ ] Evaluation specialists I can call upon for help
- [ ] Colleagues who are willing to help me with evaluation
- [ ] Publications on evaluation available to me
- [ ] Award programs for recognizing outstanding evaluation work
- [ ] Funds earmarked for program evaluation
- [ ] None
- [ ] I don't know
- [ ] Other, please specify

9 Does your supervisor assess your performance in part on whether or not you conduct program evaluation of your programs? (in other words, is "conducting program evaluations" one of the criteria on which your performance is assessed?)

- [ ] Yes
- [ ] No
- [ ] Not sure

10 By what means does your administration attempt to motivate you to conduct program evaluation? (please select all that apply)

- [ ] My administration doesn't try to motivate me to do evaluation
- [ ] Providing evaluation training
- [ ] Requiring evaluations as a condition of promotion and tenure
- [ ] Including evaluation as a criterion on annual performance appraisals
- [ ] Requiring evaluations as a condition of continued program funding
- [ ] Award programs to recognize outstanding evaluations
- [ ] Providing evaluation specialists to help me
- [ ] Including evaluation plans into state extension plans
- [ ] Using my evaluation results in program decisions
- [ ] Providing funds for program evaluation
- [ ] Talking about evaluation frequently
- [ ] Verbal "pats on the back" for conducting evaluations
- [ ] Being a role model by doing their own evaluations
- [ ] Other, please specify

11 Which one of the choices below would you consider the MOST EFFECTIVE in motivating you to do evaluation (whether your administration uses it or not)?

- I didn't select any above
- Providing evaluation training
- Requiring evaluations as a condition of promotion and tenure
- Including evaluation as a criterion on annual performance appraisals
- Requiring evaluations as a condition of continued program funding
- Award programs to recognize outstanding evaluations
- Providing evaluation specialists to help me
- Including evaluation plans into state extension plans
- Using my evaluation results in program decisions
- Providing funds for program evaluation
- Talking about evaluation frequently
- Verbal "pats on the back" for conducting evaluations
- Being a role model by doing their own evaluations
- Other, please specify

12 What is your greatest barrier (limiting factor) to conducting evaluation of your programs?

- Lack of time
- Lack of skills in program evaluation
- Lack of incentive
- Other, please specify

13 (Please use this space to give any details about the barriers you face)

14 Please indicate the extent to which you agree or disagree with each of the following statements.

	1 Strongly Agree	2 Agree	3 Neither Agree nor Disagree	4 Disagree	5 Strongly Disagree
Program evaluation is an important part of my regular work.	1	2	3	4	5
I usually drag my feet about doing program evaluations.	1	2	3	4	5
Using the results of program evaluation helps improve Extension/Landowner Assistance programs.	1	2	3	4	5
Doing program evaluation takes time away from the"real" work of Extension.	1	2	3	4	5
Performance assessments for Extension/Landowner Assistance professionals should not include criteria related to program evaluations.	1	2	3	4	5
Conducting program evaluation contributes to the meaning of my work.	1	2	3	4	5
There's really no consequences of NOT doing program evaluation in my state.	1	2	3	4	5
Extension professionals in my state are not recognized for conducting program evaluations.	1	2	3	4	5
I would conduct program evaluation even if it wasn't expected of me.	1	2	3	4	5
I rely on others in the organization to worry about program evaluation.	1	2	3	4	5

15 How frequently have you conducted each of the following activities during the past 12 months? For example, if you select "about half the time" it means that you have done the activity with about 50 percent of your programs in the past 12 months.

1 Never	2 Rarely	3 Not often	4 About half the time	5 Often	6 Most of the time	7 Always

I have asked participants for their reaction to a program at the end of an educational event (such as whether they thought the event was worthwhile).

1 2 3 4 5 6 7

I have measured changes in either knowledge, skills, attitudes, or aspirations of program participants at the end of an educational event.

1 2 3 4 5 6 7

I have re-contacted program participants some time after my program was completed to assess changes in their knowledge, skills, attitudes, or aspirations.

1 2 3 4 5 6 7

I have measured changes in behaviors or practices of program participants that have occurred as a direct result of my programs.

1 2 3 4 5 6 7

I have measured changes in long-term conditions that have occurred as a direct result of my programs.

1 2 3 4 5 6 7

16 Which one of following most closely describes how you evaluate your programs MOST OF THE TIME?

None. I don't evaluate my programs	Ask for participant reaction at end of event	Measure changes in knowledge, skills, attitudes, or aspirations at end of event	Re-contact participants to assess change	Measure changes in behaviors or practices	Measure long-term conditions
1	2	3	4	5	6

17 Which of the following sources of information on program evaluation have you sought in the past year? (Select all that apply)

- [ ] I have not sought information on program evaluation in the past year
- [ ] Contacted Extension evaluation specialists on campus
- [ ] Reference materials that I own
- [ ] Reference materials that I borrowed
- [ ] Reference materials provided for me
- [ ] Program evaluation websites
- [ ] Asked colleagues in my office for help
- [ ] Relied on expertise of others on my teams
- [ ] Other, please specify

18 How would you rate your confidence level in conducting program evaluations?

- [ ] Very confident
- [ ] Moderately confident
- [ ] Neither confident nor unconfident
- [ ] Moderately unconfident
- [ ] Very unconfident

19 For each of the following evaluation skills, please rate your desired skill level by clicking on the button that most closely reflects your opinion.

	1 I need a lot more skill	2 I need a bit more skill	3 I know enough now
Writing measurable objectives	1	2	3
Developing evaluation plans	1	2	3
Selecting evaluation methods to use	1	2	3
Conducting needs assessments	1	2	3
Conducting focus group interviews	1	2	3
Developing a survey instrument	1	2	3
Testing a survey instrument	1	2	3
Analyzing evaluation data	1	2	3
Choosing sampling techniques	1	2	3
Preparing evaluation reports	1	2	3
Using evaluation results	1	2	3

20 *In which state do you currently work?

- Alaska
- Alabama
- Arizona
- Arkansas
- California
- Colorado
- Connecticut
- Delaware
- Florida
- Georgia
- Hawaii
- Illinois
- Idaho
- Iowa
- Indiana
- Kansas
- Kentucky
- Louisiana
- Maine
- Maryland
- Massachusetts
- Michigan
- Missouri
- Mississippi
- Minnesota

- Montana
- Nevada
- Nebraska
- New Hampshire
- New Jersey
- New Mexico
- New York
- North Carolina
- North Dakota
- Ohio
- Oklahoma
- Oregon
- Pennsylvania
- Rhode Island
- South Carolina
- South Dakota
- Tennessee
- Texas
- Utah
- Vermont
- West Virginia
- Virginia
- Washington
- Wisconsin
- Wyoming
- Washington D.C.
- Place other than above

21 *Which of the following most closely describes your job classification?

- Extension agent with county/regional responsibilities
- Extension specialist with state-wide responsibilities
- Extension program assistant
- State Forestry Department employee
- Other -- with some extension responsibilities

22 *What is your highest level of education?

- High school diploma
- Technical degree
- Bachelor's degree
- Master's degree
- PhD
- Other, please specify

23 If you work for a land grant university, how many years have you served in the Extension system? If you are not employed by a land grant university, please estimate the number of years you have done extension-related work.

24 Do you have any recommendations to improve the practice of program evaluation in your organization?

25 What is your age?

26 In some states, natural resource extension agents are considered "faculty members," and in other states they are considered "professional staff." If you are employed by a Land Grant University, please select one of the following that best describes your position. If you are not employed by a university, please skip to Question 28.

- Tenure-track faculty
- Non tenure-track faculty
- Professional staff
- Not sure
- Other, please specify

27 If your position is considered tenure-track faculty, have you achieved tenure yet?

- My position is not tenure-track faculty
- Yes, I have achieved tenure
- No, I have not yet achieved tenure
- Not sure

28 What percent (%) of the funds for your salary currently comes from grant ("soft") dollars?

29 What is your job title?

30 Do you have any additional comments?

Thank you! This is the end of the survey. Please click "submit" to complete the survey and proceed to the optional "draw for prizes."

SUBMIT

APPENDIX B NON-RESPONSE EMAIL AND QUESTIONNAIRE

Dear _____,

You may remember receiving an email invitation to participate in a national on-line survey regarding Program Evaluation in natural resources Extension a few months ago.

I am a PhD candidate at the Univ. of British Columbia and an Extension Specialist here in Canada and am doing a study about US Extension agent/specialist attitudes and behaviors regarding program evaluation (..being a former county agent myself).

I am writing a small random sample of people who weren't able to respond-- to ask five short questions to ensure that there are not any major differences between those who responded to the survey and those who weren't able to respond that might bias the results.

I am writing to see if you wouldn't mind taking about 1 minute to answer a few questions in a quick reply email to me. You can write your answers below. Your answers will not be shared with anyone.

1) Are you campus-based (as opposed to county/region based)? (yes or no)

2) Is your position a "tenure-track" position? (yes, no, not sure)

3) If your position is "tenure track," have you achieved tenure yet? (yes, no, not sure) (Leave blank if your position isn't tenure-track)

4) What percent of your salary comes from "soft" (grant) funds?

5) About how many years have you worked in Extension or Extension-related work?

That's it. Thank you so very much. Please respond by this Friday, April 2 or as soon after as possible the following week.

If you are interested in learning more about the study about evaluation behaviours and attitudes of Extension professionals, please let me know and I will ensure that you get a summary of the results. I will also be presenting my results at the Association of Nat. Resource Extension Professionals meeting in W. Virginia in May.

Thank you again.!!

Shawn Morford
PhD candidate
Faculty of Forestry
University of British Columbia, Canada
Tel: 250-363-0769

APPENDIX C BARRIERS FACED BY RESPONDENTS

(Answers to Survey Question 13)

Lack of Skills

- I am not a sociologist and so these types of questions I have no formal schooling in.

- I'm not an evaluation specialist, so I don't know immediately what to do beyond the obvious and traditional methods.

- Lack of skills based on the organizations not providing leadership, support, and professional development activities.

- Not feeling absolutely comfortable in composing questions

- Methods for effective evaluation are not provided and perhaps not known.

- Don't really know what tools would be best for which program

- Our evaluation "specialist" has limited knowledge regarding statistical design of evaluation instruments.

- Don't have statistical background, have no experience in designing surveys or interpreting results.

- Most of us working in county programs are Masters level. We don't have a deep well of scholarly evaluation methods from which to draw.

- Lack of skills also a barrier

- If evaluations were required in the programs I am involved, the only barrier I would have is how to do them (and follow through) effectively.

- Lack of skills is also important. If I had more time, I could study evaluation methods more.

- I don't have either time or knowledge to do much more than the surveys that are provided for me to use. I would like to know, however, how effective the training was over time, and so a follow up survey would be useful.

- I perform evaluations of my programs, but am not sure how effective they are since all I know about program evaluation is what I learned in classes.

- training needs to be evaluated and offered.

- As a relatively new Extension educator, lack of training in program evaluation is also a barrier

- Basic training beyond just surveys or specific program evaluations is lacking, and if you are interested, you have to pursue it on your own

- While training is provided it is infrequent and help never seems to be there when you need it

Time prioritization issues

- I know I need to do more evaluation; the question is time!

- It's not that evaluation is bad it's that it takes a lot of time that really busy Extension people don't have time to do.

- Insufficient time to conduct evaluations due to staff cutbacks

- Lack of time is also a barrier.

- On the lack of time: program evaluation is a long-term project that always seems to take a back seat to today's emergencies.

- I'd say it's a combination of lack of time and lack of incentive. There's a point of diminishing returns -- you could spend every day evaluating every aspect of the program. So you need to choose efficient methods that provide the most important feedback. More time would allow more feedback.

- One must build time for evaluations into any program, in order to conduct and evaluate them.

- Too many other duties to deal with. Once a project ends, it's on to another.

- Not enough time in planning for evaluation.

- In the past two years my job has gone from that of assisting 3 state specialists to covering for two state specialists that have left and assisting the third. So there is no time for in depth and meaningful evaluation.

- No time

- Dealing with too many programs to effectively manage them or evaluate them

- Unfortunately, the job is so busy that we often move from program to program with little to no time to "debrief" or consider effectiveness of a program and/ or how it can be improved.

- I have many program initiatives going on at the same time and it is hard to find time to conduct post-evaluations when my energies are on new initiatives.

- Taking time for more thorough evaluations would interfere with productivity

- Lack of time to conduct post program evaluations.

- My program is spread so thinly among many clientele that there is little time for thorough evaluation.

- More things (projects, programs, and consultations) to do than time required to do an effective job of evaluation. Program leader has indicated that 60% of my time should be on programming and 40% on planning and evaluation. Just too many jobs required of specialists and too much demand from county agents to spend that much time in evaluation.

- Mostly youth programs done in limited increments of time. Evaluation would take up too much teaching time and/or typical evaluation format not appropriate

- I need assistance in being able to gauge what is appropriate. I see many folks doing end of program type evals which give bad information or misleading info.--not real evaluation information.

- It is hard to find time for 6 month, or long term follow up evals when you are onto delivering the next programs.

- Adequate program evaluation is very time consuming. Proper write-up and statistical analysis take a tremendous amount of time, time that could be used to prepare for future activities/programs.

- I work 60 hrs a week as is, there is precious little time for everything.

- Formal program evaluation techniques take quite a bit of time and planning. I am focused more on program content to the detriment of focus on evaluation.

- I am personally motivated to do continuous improvement, but the scope and complexity of my day-to-day responsibilities limits my ability to devote much time to evaluation.

- Time

- Too many expectations, too few resources

- My greatest problem is in prioritizing programs. However, we are certainly graded on the impacts of our activities.

Lack of Funding

- Program evaluations are a useful tool to refocus agency programming, but are not regularly used due to staff and budget constraints

- Lack of money to pay to do surveys well or hire someone

- Adequate funding for testing control groups to compare to our attendees. Lack of ability to compare our program attendees to our audience. No funds available for follow-up to determine if program actually changed behavior. We rely on outside grants to publicize our programming via mail and for travel expenses for presenters. -There are little funds available in-house, due to the ongoing budget crisis of our entire state

- Insufficient time to conduct evaluations due to staff cutbacks

- Lack of operating budgets (for postage, etc)

- Another challenge is limited funding

- Funding to conduct such evaluations is limited also.

- Constantly asked to do more with less time and money

- Budget crunches

- I am soft money funded and evaluation is not funded

-Good evaluations can be very expensive. Our resources are constantly being cut but we are still expected to evaluate.

- Too many different programs - feel all alone in this.

- Increasing numbers of educational programs expected with decreasing staffs

- Staff and budget cuts and increasing workloads leads to things being dropped. Evaluation sometimes gets dropped

None

- I don't really think that there are any prohibitive barriers. The ability to evaluate program performance allows for program improvement and increased personal satisfaction in doing a good job

Methodological difficulties

-I live and work in a small community so there are limited opportunities to evaluate my clientele without them running away. I believe Participatory Action Research is an effective way to design and conduct our programs.

- In dealing with natural resource issues, many of the results I achieve are many years down the road, and/or serendipitous. It is difficult to measure beyond knowledge gained at the end of a workshop.

- Process needs to be quick and easy to do.

- It is relatively easy to measure if an audience feels a program is successful and increases their knowledge. It is more difficult, time consuming and costly to determine the true economic impact of programs.

- I think written evaluations are tough in that I believe people often give you what you want to hear.

- The forms provided operate much as course evaluations, so prior paperwork and form estimates require lead time. The system also requires volunteers to administer the formal evaluation and seal-mail back to source department. It is a cumbersome structure for many of my audiences to get realistic feedback.

- Lack of response to surveys

- I am sure they are effective, but am not sure if they are the types of evaluations that would be the most beneficial to my unit

- In many cases, the program is not conducive to follow up (logistics of contacting people) or that it may be that you can evaluate whether people liked the program or had an increase of knowledge but it is much more difficult to determine if their behaviors have changed

- It is hard to collect information that provides concrete evaluation or impact of Extension programs. Also, as I try to do this, I think I often underestimate the value of these programs by trying too hard to provide "solid" data with regard to behavior change, etc.

- Hard to come up with evaluation instruments that reflect the true knowledge gains of your clientele

- I think 'evaluation' forms (e.g., customer satisfaction surveys) in general are so commonplace that clientele ignore them. That is, we also have issues with incentives for our clientele to respond

- I would enjoy a discussion or training on methods to perform effective and informative evaluations

- Compiling the data collected into a coherent interpreted usable product -- I call it the "yes, but what does it mean?" syndrome

- Also, it is sometimes difficult to get the needed addresses, emails, etc. About program attendees from county educators

- Clients tired of surveys

- Many of my programs seek to motivate improvements in practice or understanding of concepts. Detecting these changes in my clientele group is challenging

- Although I recognize the value of evaluation, especially to see if my programs make a difference, there continues to be more and more red tape. At Oregon State University, for example, if we (Extension Educators) plan on using the results of our surveys in public presentations or journal articles, we must go through a formal review by the Institutional Review Board (IRB) for "human subjects research"

- It's more that the audiences don't like to do or pay attention on how to fill it out and give you bad scores when everyone else gave you good ones but it draws down your total. This type of thing does not get reported you just see a mean.

Lack of an organizational system for evaluation

- We currently have no official evaluation system for our programs

- The organization is aware and wanting us to do more significant evaluations (more explicit impacts) but not yet giving us all the means to make the transition

Lack of rewards for doing evaluation/lack of administrative incentives

- no strong incentive to do, but no incentive not to either

- We are becoming so grant driven that other than doing evaluation as called for by the terms of the grants.

- my tenure home could care less about evaluation (but Extension does provide resources; hence the dual answer).

- Our evaluations are sent to the university and we never see the results from what they do with them. My program evaluations from the participants have always been high but it does not make any differences on evaluation or anything else. Brown nosing goes a lot further for promotions

- The tenure system works against evaluation. There is little incentive to complete essentially "research" that is outside my expertise for which I can be rewarded.

- No personal or office benefits to making time/effort for evaluations

- My organization probably assumes that I do program evaluation - but it's actually pretty random

- Although it would be easy to say "lack of time" is the greatest barrier, if there were a major incentive - e.g., considered in performance appraisal - then I would find the time!

- Evaluation is easy to leave behind, since the reward for doing it is not so great

- The expectations to constantly produce and perform to a certain level do not allow enough time to properly evaluate a past program

- Our reward system is to make sure that we are servicing requests from the county agent network and providing effective information in natural resources to these folks. Program evaluation would be 3rd or 4th on the list

- I do end of program surveys for every workshop that I do, but it is mandated by the national program that I use, rather than my agency

- Lack of skills based on the organizations not providing leadership, support, and professional development activities

- My organization gives a lot of "lip service" to evaluation but does not devote the resources to it

- Cooperative programs with other agencies/groups don't see evaluations as important

- My primary barrier is there is no incentive to improve generated from the organization I work for. The pay is less every year. My motivation is self induced because I want to do the best I can and work toward improvement. It simply makes no sense to develop a training program and then have no evaluation to know if it is helping someone

- There is no time for detailed evaluation. If it's to be done, I either need someone to do it at my direction (an evaluation specialist; the carrot approach), or they need to make it important to me by, for example, requiring a major program evaluation before you get tenured or promoted (Nebraska did this while I was there); i.e. The stick approach

- Ongoing rut of prep -> program -> cleanup ->prep -> program -> cleanup and evaluation seems to be a sidetrack

- Also lack of incentive--we're told we need to do some evaluation, but no specifics

- New things are added to the plate while nothing is taken off. Evaluations take time and they are need there is little recognition of that and little effort to give help to get it done

- Requires dedicated funds and person hours to plan, execute, analyze, report, but not funded because it is often viewed as "non-productive"

- Local County pays all my salary so that I am not financially rewarded for doing good work

- No one cares

- Currently, in Natural Resources Extension, little recognition of programs exists by county or area agents who do a good job evaluating programs

- There do not seem to be any negative consequences if I do not evaluate programs

-Two different administrators with differing views as to the value of evaluation

- Because I am federally funded, my evaluation results are only used in annual reports. I do not qualify for promotion in the Extension system - less of an incentive to go the extra step of evaluating

- I sense that we are expected to "Do More" versus perhaps doing less but more completely

- Quality evaluation takes time and money. I have seen little evidence that this is a real priority for my organization in that time and money are not earmarked for evaluation. It's expected and required, but not well supported

- Evaluations are a useful tool to the educators but do not always reflect an educators worth. Thus evaluations that may be most useful to the educator will be worded differently than evaluations that are used for promotion and tenure. Any results can be modified by the way questions are asked. Not all programs warrant evaluations, and the purpose of the evaluation needs to be clearly stated to the audience, the educator and administrator

Scepticism about value of formal evaluation

-I am also sceptical of essentially self-evaluation. It would seem that someone else should evaluate my work that can both shield respondents and faculty. Results of external peer-reviews (as is done for much on-campus teaching) could be motivated in tenure decisions as "demonstrated performance and improvement." Even with the best methods and intentions, are self-administered evaluations credible?

- my plate is full and feedback is generally positive to my programs so evaluation slips

- some programs don't lend themselves to evaluation. For example a day long field program for loggers

- I also believe that in evaluating our own programs a bias in reporting results. The foxes should not watch the hen house.

- Evaluations of educational presentations I give at conferences and in classes are often done by those facilitating the events, and I consider them sufficient, for the most part, for me to gauge my effectiveness without my having to devise my own evaluations

- So many of the evaluations that I have seen other specialists do seem valueless

- Formal program evaluations conducted in the past did not produce changes in program delivery or content

- I work with a small stable mature audience that has filled out many evaluations over the last 30 years. Their advice is better than forms

- Evaluation is a hindrance to what I consider are the more important parts of my job (field assistance, study, program preparation, etc.)

- I suspect that some folks are too nice to critically evaluate a program

- I judge by the questions and interest on-site. I also view return invitations and word of mouth as a better gauge on the presentation

- Follow-up on change of behavior is not really part of my section of programming, but handled by the organizers of the conferences. Also CEU crediting for my sessions serves as a "value judgment" on the topic and credential for information transfer by the program and sponsor groups. I find most people give good reviews unless they have an issue. Bubbles don't give much info, and I rarely get text feedback to make critical self-evaluations

- I'm just glad to organize, facilitate and pull off good programs, and if the program is good the people will verbally tell you, and if it sucks they won't be at the next one. Using Extension committees to identify the needs of the community really helps in the success of

programs, If you listen, build a quality program, workshop, or presentation, People will come and they will tell their friends and family, and more will attend the next one. Some folks think filling a questionnaire at the end of a meeting is waste of time

- Whenever I find something is not working or is particularly effective, I do respond and try to build or change as necessary. Most of the Extension colleagues I know take very tactile, hands-on approaches to their programs. Conducting program evaluation as the bean-counter administrators design is just a weak suit for many Extension professionals. If it were, we would probably be doing some other kind of work than Extension

- Disinterest is the only barrier we have towards program evaluation.

- Change comes slowly. Evaluation after a meeting (happy face evaluation) is of limited use - mainly for meal, room, and general subject choice or immediate feedback on the quality of the presentations

- Motivation

- Educators give people a foundation upon which they can build or choose to do nothing. The idea that we are responsible for our client's actions is politically driven and in some ways unrealistic

- With short programs it is difficult and not useful to have extensive evaluations

- I don't think there would be barriers

Other comments

- It is hard to choose between lack of time, lack of incentive and lack of skill. The three really are similar in importance, and interact. Always time to do one, but costs more to do more comprehensive evaluation

- Our administration has adopted the Logic Model for program development and evaluation.....Impacts vs. Outcomes. I like this model! The model has been in place for two years now and we have not implemented or adapted enough of our organizational culture and materials to reflect this change

- Many programs to manage, high volume of requests, limited staffing, limited funding

- I believe that evaluation should be nearly standardized and efficient. Taking the time to individually evaluate every workshop or training means that I would have to do this 50-60 times a year. That would be incredibly inefficient. Much of the follow-up can be a standard mailed postcard-type survey. 4-5 questions would answer much of what we need to know. This is, however, even at less than 30 cents a unit and at most a minute of time (formatting, data entry and analysis), costly in time and postage. Support of this in money and dedicated professional/support staff time would be great

- We have only 3.5 FTEs to do education for 40,000 forest owners, 2,000 loggers, and 500 natural resource professionals. We are often conducing programs in fairly broad conceptual themes (e.g., basic forest ecology), rather than discrete technological innovations. Therefore our approach leans to being extensive and intuitive, rather than being intensive and scientific, to allow more time for programming

- I value feedback. I am looking for ways to get better.

- Different Evaluation Criteria (no consistency) Among Project/Program Funding Resources at County, State, and Federal Level. Parceling out Results among Different Programs and Calendar and Fiscal Budgeting Years

- At this point I am concerned about how we define "Program Evaluation". Surely I am evaluated my programming constantly

APPENDIX D LIST OF APPROACHES THAT RESPONDENTS USED TO CONDUCT EVALUATION BEYOND END-OF-EVENT QUESTIONNAIRE

(Answers to Survey Question 4a)

Needs assessments

- Woodlot owner survey in preparation for an Extension program

- Various needs assessment activities. Regular meetings with committees, irregular meetings with stake holders in a formal setting, informal meetings with individuals, and local contacts with industry leaders.

Measuring changes in knowledge, attitudes, skills, and aspirations (KASAs)

- Pretest and post test of terms

- Pre and Post testing of knowledge gained

- We are developing 'scenarios' to use with a small sub-set of audience in conjunction with observations and questionnaires, and we are analysing historical records on our audience (license numbers kept over last 11 years)

- A simple before and after "exam" (same questions) rating their knowledge before and after the program

- At breaks between speakers, had participants form groups, list things they learned, and prioritize questions/things they still want to know

- Performed a survey of past years workshop participants on the use of video in technology transfer for Forest Stewardship Workshops

Measuring practice/behaviour changes

- I use follow-up surveys 4 to 6 months after a program to determine behavior and/or attitude changes

- Evaluate practices being implemented on the ground

- Personal or phone contact after six months to a year to see if program participants changed their farming practices

- Coverts program evaluation, collecting data on outreach and management on participant's property

- Follow-up interviews or questionnaires looking at application and retention

- We went back to a group a year later to see what they really put on the land

- Participants diagrammed their streamside area before the class, made modifications during the class. Both were photographed, as will a 6 month follow-up

- One-on-one follow-ups, documentation of forest management plans completed, documentation of on-the-ground management activities, local changes in regulations, land protection activities, documentation of volunteer hours, etc.

- Email survey of program participants whether they have planted plants we have been promoting

- 6-month post program survey to determine adoption of practices

- I have a survey that is filled out by participant when they pick up a radon test from our office. I also, evaluate quantities of materials collected at various recycling programs

- Mail surveys to ascertain whether practices have been implemented by landowners who attended meetings or workshops. Surveys are on an annual basis

- Mail surveys following courses to evaluate implementation of information presented

- I just did a post-program questionnaire that will be followed up in 6 months to look for longer term changes

- In the summer of 2003, I had a student worker conduct follow-up surveys of landowners who participated in five separate Extension forestry/wildlife educational events that were held over the course of three years. I gathered information on practices that had been adopted as a result of the events

- An in-depth evaluation of a visualization tool for stand management. The evaluation involved a quasi-experimental design evaluating pre-program, 1/2 program and full-program impacts on knowledge and attitudes as well as the educational tool and a 60-90 day follow up survey

- Compilation of audience data that hired a licensed forester, had a stewardship plan and prepared and/or harvested timber

- Sent 6 month follow up evaluations asking participants which practices/topics they learned about had actually been put into practice

- I collect names and e-mail addresses from website visitors when they download a software program on my website. I contacted those folks and asked how they used the program, how it could be improved, etc

- Phone calls to workshop participants one year after attending workshop asking about behavior change

- Program assistant visited farms with questionnaire

Higher level impacts

- In addition to standard demographics, I ask them to place a dollar value on the knowledge gained from the meeting as in "Will this save you money in the future or earn you money in the future?" Also ask for acres impacted

- Measured Cost Reduction, Improved Utilization, Jobs Creation, and Training Development

Informal evaluation (consultation, advisory committees, general observation)

- I have an advisory council of 7 individuals that meet to discuss and evaluate programming

- I maintain a contact management database for all "students". A summary of all conversations is recorded. Since I am always looking for change in skills, behavior or attitudes I usually ask for these types of changes as a part of any conversation

- Including audience members in program planning efforts in order to assess current program value and make improvements based on audience needs

- Annual Advisory Council meeting. The Council is made up of a variety of natural resource individuals from across the state

- I am constantly doing informal survey

- Discussions with individual participants

Other methods

- Follow up emails 4 months after educational event to document any longer term impacts of the training

- Participants were asked to fill out questionnaires several months after an event

- Follow-up surveys to participants 6-12 months after event. Pre-tests or initial surveys to program participants. Focus groups of past program participants

- Surveys of client groups

- Phone surveys, email surveys, mailed surveys

- Major 5 day invited review of a major program effort

- Follow up surveys several months after the event has passed

- Certain programs culminate in a community development or other project result. The extent of achieving that result measures program success in some ways

- Formative evaluation of an on-going conflict resolution program

- We use a standard "How did we do" survey for three of the classes and take demographic information on the fourth

- I have used several web-based surveys as well as mail surveys approximately six months after programs. I've also used a web-based survey to document impacts from a web page

- Focus group discussions

-Informal one-on-one discussions between sessions or at other events

- Use participant listserv to gather input

- Two follow up questionnaires, two months after the event

- We work with a team of teachers who review our program materials. We have also conducted some interviews with selected teachers on the effectiveness of some of our programs

- Follow-up survey of information use

- Follow up surveys 6-12 months later

- Focus groups

- We did a 6 month follow-up of a backyard wildlife landscaping program and a bird-monitoring program

- Follow-up phone calls to sample of participants in Citizen Planner program

- I use a written evaluation tool for all programs where at least ten landowners attend. For major programs such as "Master Tree Farmer," we do one year post-program evaluations for all participants

- Pilot testing of all programs

- A telephone survey was carried out with participants of a school program about 1 year after involvement to determine impact

- Follow-up personal contact with clientele to see if information was useful

- I have done mail out surveys 1 year following a program or event to see if any practices have been implemented

- We have developed an on-line survey for home landscape practices

- USDA Forest Service evaluated our PA state-wide urban and community forestry program through site visits, in addition to the required annual reporting. The Extension Service also requires evaluations of our Plan of Work, using procedures that we design within general guidelines

- I had them to do both, fill out a questionnaire at the end of a meeting, as well as a questionnaire about Extension and upcoming program requests

- In one case, a formal survey to see the impact, determine future needs and determine businesses developed or improved based on the programming

- Mail questionnaires when contacts are made requesting program information

- We haven't actually done it yet but we have been planning for a survey (in 2004) of 10 years worth of people who have taken our forestry short course

- Post-program mail survey

- I have conducted surveys of pesticide use among Christmas tree growers. This has helped me evaluate programs in the past. The last big survey was in 2001. We are also holding a meeting this week with county Extension agents to talk about evaluating programs and to develop surveys for after meetings. This survey requested information for the last 12 months, but next year, I'm sure my responses will be a lot different.

- Local agents generally plan and carry out evaluations of programs in which my participation is included. I do not always see the results of the evaluation but the agents continue to call on me for participation.

APPENDIX E COMMENTS AND RECOMMENDATIONS FROM RESPONDENTS

On the value of evaluation

- I don't like doing evaluation, but I value the outcomes of it.

- "Evaluation is becoming a bigger part of our programs. We are being asked to show our worth more to funding agencies."

- Most of Extension's problems with state budgets can be traced to not doing program evaluations program wide AND not using program evaluations to market our programs at the state, regional and national levels!

- I'm not a big fan of evaluation, but I understand its importance in determining program effectiveness (and therefore, continued funding).

- Most of the official evaluations are done at the state level rather than the district level. The "evaluations" I do are mostly very informal. I can see the difference in attitude and forest management practices after working and talking with a person for 3 weeks, 5 years, etc., but most of the time I do not have an evaluation form that I am filling out and submitting.

- "Measuring" changes in attitude or practice is a correlative analysis. There is a great risk of misinterpretation into cause- and-effect of this type of evaluation or survey work. Evaluations are great for getting audience gut reactions to something they feel strongly about. A negative reaction is more likely to illicit a response than a positive reaction. Evaluation tools have a value if there is no threat of them being misused for personal advancement.

- Some of the answers/comments are a result of this position being new and in a remote area of Alaska. I have been in this position for only 5 months and have not been exposed to all of the particulars of the organization. I personally feel that the evaluation process is very important and have established relationships with "stakeholders". Local residents whom will evaluate and critique program/results.

- Colorado State Forest Service is a branch of Colorado State University. Therefore we are part of a Land Grand University but also a service agency. Our primary concern is acres treated, not measuring change in opinions. Educational efforts are determined successful when practices are implemented on the ground.

- Program evaluation usually lies with our staff in Ft. Collins

- I have the luxury of having limited Extension appointment. I think many have an inordinate amount of accounting for hours and activity which demonstrates a need for accountability operating in a climate of a) mistrust on the administrative level and b) administrative need to look at numbers rather than value to justify funding allocation to meet a very important mission for Land Grant institutions.

- I think we should do more internal evaluations. As a professional (22 years in the work place and 5 in academia), it is evident that the leadership and many of the Extension specialists have lost touch with the people producing.

- Effective evaluations are imperative to knowing if you are helping and if you need to improve.

- I hope your survey helps people understand that Extension has got to produce results that benefit the tax payer and producers (economic development). The money is running out and the programs are suffering.

- We are doing evaluations. External factors are important in choosing to do them, but also in influencing type.

- We were just starting to look at long-term evaluation of our program

- Evaluation is key to judging how well you have done in impacting your audience. The best time to develop the evaluation is during the planning stages, so that the program/activity will address the issues you want clientele to learn.

Recommendations

Administrative incentives/rewards/prioritization for evaluation

- "Use a carrot, rather than a stick"

- Encourage agents with positive reinforcement, i.e., promote, give merit raises and publicly recognize those who are doing a good job!

- It has to be part of tenure and promotion

- Make it a priority

- Positive program evaluations should have positive rewards

- Make it part of performance review

- Provide incentive such as recognition and monetary rewards

- Awards, rewards, recognition for those who conduct evaluation -- not just for those who are good agents

- I'd like to hear it said, "It's better to do less but do it better, than it is to do more" repeated at every Admin update

- My organization verbally encourages program evaluation. However there is little incentive for agents that are not (doing evaluation).

- Negative consequences for neglecting evaluation - reduced raise, tar and feathering, etc.

- Enforce a stronger system of rewards and requirements

- Emphasize the importance of evaluations to be sure we are satisfying stakeholder needs

- There should be an evaluation requirement of all programs before approved by Program Leaders

- Need to prioritize programs and have administration support when certain programs are dropped. We are still at the point where we try to be all things to all people. In this mode, it is difficult to concentrate on the long-term, sustainable programming and judging your impacts

- Keep evaluations as tools for educator self assessment and only use for performance evaluation when the need arises (job related grievance issues)

- Make it mandatory, and tie them to performance evaluations

- Make it a part of annual performance evaluations

- I do evaluations to help make my programs better. I have been doing this a long time and, in general, I know from observation if a program is effective or not. I believe that much

time and effort can be wasted on evaluations if there is no purpose except "you should do evaluations". That is why I believe explaining how and why evaluations will be used is important. Often administrators can't tell you this other than it is required.

- We need success stories showing how good evaluation strengthens programs. At present evaluation is a "must do" instead of a "want to do activity."

Standardization/establish evaluation system

- Streamline the system- we NEED evaluation; however, we are paid to provide a service to the public. Much of what we need can be answered in 4-5 questions. Did you establish a business or alter your business or management of you property based on this training? How? Do you feel that you learned information that was useful to you? Would you recommend this workshop? Or questions that are appropriate to the audience

- We need to adopt a policy for doing so

- Develop a program evaluation system for research also

- We would use an evaluation tool that was simple, easy to implement and interpret

- Standardize the practice at both the University level and the county level. Possibly put one person or a team of persons in charge of the program for a set period of time, say 3 years

- Give specific guidelines/trainings as to how many and type of program evaluations are expected per year. This is especially important for newly hired educators

- Integrate methods for evaluation into training on all subjects so it does not seem so daunting or mysterious 'on its own'

- Consistent evaluation tools across organization for statewide programs. These are provided with some programs, but not others

- Develop generic evaluation process to use with multiple programs/projects of agency

- Consistent Evaluation Criteria. Make Evaluation a Part of All Extension Programs, not Just the New Program Initiatives

- Identify the program evaluation tools that Extension professionals find useful and easy-to-use that give the administrators the information they are after. If this were tackled and made available to the Extension agents, we may be able to resolve the issue

- Provide standards

- Make them user friendly. The results should be available and to the lowest in put level

- Evaluation should be built into the program planning process

- There needs to be a better linkage between programs and evaluation. In other words, for every program there should be an evaluation strategy on the web page

- Make evaluation part of departmental and Extension program discussions for everyone, not just something that comes up before you apply for tenure and not thereafter

- Budget time and resources in work plans for evaluation - practice, training and professional development

Provide evaluation expertise/resources

- A reference notebook or set of resources would be helpful, at least one set per office.

- This past year I have asked our campus evaluation specialist for help for an evaluation and did not receive a reply.

- A web site would be helpful.

- Have a good suite of instruments handy

- Provide an evaluation service section that evaluates programs for all specialists and writes up the results for submission to funding agencies

- Provide us with the evaluation instrument and tell us exactly what they want; otherwise it is guesswork, and quite frankly strongly biased

- Provide state level professional staff focused program evaluation that will provide leadership, training, and support for this activity; require it for annual performance appraisals

- Have a dedicated evaluation specialist that can conduct in-service training and serve as a resource

- More assistance in developing evaluation methods

- Providing resource to conduct meaningful evaluations

- Have an evaluation specialist available to assist with and/or do the evaluations of long-term change

- Access to a statistician

- Provide a specialist

- Enough specialists for assistance.

- The ideal situation, in my mind, would be to have an "evaluations office" - a group of evaluation specialists that would work with Extension faculty to design, administer, analyze, and summarize evaluations (both short- and long-term). This would clearly lighten the burden

- If it really is important to the organization, the organization should provide the expertise to conduct evaluations

- Putting together a packet of pertinent evaluation literature for natural resource Extension educators

- I would still make program evaluation optional but provide convenient materials for doing and interpreting evaluations on a regular basis

- Make it a specific job duty for central office staff to assist field personnel

- Evaluation omnibus men to assist others get evaluations done

- Provide resources to assist agents

- Provide more resources. Need people to help us develop targeted, efficient eval plans that take us beyond traditional and obvious methods. I want to focus on content, not on developing innovative evaluation techniques

- Don't think that a person advising us on evaluation importance and techniques will help much. Real help would be having an evaluation specialist you could go to every couple of years who would help you actually conduct an in-depth evaluation -- their job would be to help design a survey or other materials, etc. And to help administer it

- Hire more program evaluation specialists

- Make available adequately trained evaluation specialists and staff resources to conduct evaluations

Mandatory Training

- I suggest that all new Extension agents be required to take a one-day (minimum) training in program evaluation, and one-day training program planning.

- Mandatory training for all new Extension agents and specialists

Optional Training

- Offer regular trainings

- Training on methods and use of materials must be provided

- Increased training opportunities on methods

- Offer in-service training in program evaluation

- Additional evaluation workshops

- Lots more meaningful in-service at the program level. Large group is useless

- The lack of time to do in-depth evaluation development and reporting is a huge factor

- We need to do follow up evaluations, to determine actual behavioral changes

- Evaluation should be taught to new Extension educators

- Training in database skills or resources to provide expertise in this area.

- Training

- more funding for more education

- Training

- Provide annual training opportunities to enhance evaluation potential

- Provide training

- In-service training

- In-service training for specialists on evaluation of programming

- We need more skill development in good, practical, applicable evaluation methods. This must be more involved than a two hour lecture on how to do a particular method.

- Additional trainings on program evaluation.

- Provide more training

- Effective training sessions.

- Provide training

- Readily available training in program evaluation is needed!

- Training in statistical soft ware

- Training to perform evaluations

- Self paced, distance delivered, user friendly information / training resources on program evaluation

- Common-sense based, concise information from which to work from. I'm not interested in talking about evaluation philosophy all day (which we have done) -- how about some practical tips and templates to work from. Should be accomplishable in a 2-hour training period.

- Provide regular training opportunities and an updatable handbook that includes many examples with "templates" to use in various situations

- Get research (evaluation) through University departments (Natural Resources, Survey Center) for major program efforts (for instance: Do landowners with stewardship plans implement recommendations "better" than non-plan owners? Do they use a forester? Is better silviculture applied? Etc...

- Conduct in-service training or provide assistance in developing evaluation tools; make it a priority and use the information gained in planning for the next year.

Funding dedicated for evaluation

- More funding for more education and support

- Provide funding to do some of the more expensive surveying such as mail surveys, phone surveys or personal interviews

- Evaluation is central to what we do; however, it costs in resources and time. Most of the evaluation I do relates directly to funded "research" and Extension. When this is the case, we do evaluation. Otherwise, it is not as high on the list. Nonetheless, with support funding, which would purchase time to do evaluation, we would do more

- Provide funds

- Funding support for evaluation

- Provide targeted funding

- Include results of evaluation in planning and communicate that this is being done

- There may be some sort of evaluation done at the State Office level, but the results don't really get to the field. This should change

- Show examples of how a program evaluation substantively improved a program or increased support for the program

Other comments/ideas

- Not time consuming. Keep it to a quick simple task. The shorter, and faster, the more often it is done.

- Evaluate all various types of activities

- More sharing of tools - examples from others with similar programs

- Develop a newsletter and feature successful program evaluation or evaluations each issue. Give examples of people in our organization and what they are doing relative to evaluation and focus on what makes their evaluation successful

- It is an evolving process that hopefully becomes stronger with time and interaction with other specialists

- I need to specialize so that I'm not programming in so many different disciplines.

- I think gaining more information on the long-term impact of our programs would be very valuable

- Just do it!

- Let those of us who know our audience best decide how to evaluate

- None- it is done well...

- Don't ask....

- It's a new position in a region the size of the Virginias, which has never had an Extension presence

- Allow more time for this and perhaps hire people who can be

- The impact of Extension programming needs to be more quantifiable. How much good are we doing?

- Look at the Logic Model

Revised
8/11/03

OREGON STATE UNIVERSITY EXTENSION SERVICE

Periodic Faculty Review and Professional Development Plan

Instructions: To complete biographical data section, move cursor to data line, engage insert key, and type in data. Repeat for each data line. To put a check in a data box, move cursor to lower, right-hand corner of box and left click mouse.

Faculty Member:_____ Evaluation Period:_____

Current Rank:_____ Years in current rank:_____

FTE: _____ Tenure Status: _____

Source of funding (check all that apply): ☐ federal ☐ state ☐ grants ☐ county/district ☐ other

Nature of assignment (check all that apply): ☐ single county ☐ area* ☐ multi-county** ☐ statewide ☐ multi-program

List assigned program area(s):_____

Number of years in current assignment:_____

Job Description Reviewed? ☐ Yes ☐ No Job Description Current? ☐ Yes ☐ No Last Revised_____

Following documents reviewed: ☐ Current Vita ☐ Plan of Work ☐ Report of Accomplishment (including Contact Report)

Conferred with Program Leader(s) to complete evaluation? ☐ Yes ☐ No

Conferred with Academic Home Administrator to complete evaluation? ☐ Yes ☐ No

Reviewed input from multi-raters to complete evaluation? ☐ Yes ☐ No

Area assignment: Program assignment includes more than one county, but program is delivered without regard to percentage of faculty member's time spent in each county of the assignment.

Multi-county assignment: Program assignment includes more than one county and there is written expectation in the position description that a certain percentage of the faculty member's time be spent in each county of the assignment.

Signature of Immediate Supervisor Date	Signature of Program Leader Date
Signature of Faculty Member Date	Signature of Program Leader Date
Signature of Department Head Date	Signature of Extension Administrator Date

Signature of all administrators indicates concurrence with the review. Signature of faculty member being reviewed indicates they have read the review and discussed it with their immediate supervisor. Space on page 8 furnishes the faculty with an opportunity for written comments, explanation or a rebuttal to the evaluations placed in the faculty member's personnel records file. Disagreements on the contents of the file should be handled through normal University appeal procedures.

Instructions: Objective, constructive feedback is essential to the development of faculty and team productivity and performance. Using this instrument as a guide, discuss each performance dimension, including the performance indicators listed, the employee's demonstrated skill and ability in that area, and any suggestions for development. **Going from top to bottom, the performance indicators are presented as progressive levels of achievement.** Check the performance indicators the faculty member has achieved in this evaluation period. (To make a check, point cursor on lower, right-hand corner of box and left click mouse.) Depending on the job description, "not applicable" may be appropriate for some performance dimensions or indicators. Comments should relate to the performance indicators, identifying specific strengths and areas for improvement. Comments are to be based on the current job description and the faculty member's length of service, rank and experience in the current assignment. Professional goals listed for each performance dimension should be jointly agreed upon by the faculty member and the immediate supervisor.

Teaching, Facilitation, and Other Assignments Extension faculty members are expected to include appropriate educational programming for both youth and adults within the scope of their job responsibilities.

1. Program Development

Program Planning:

☐ Demonstrates ability to recognize, understand and facilitate opportunities and to coordinate the necessary resources that best respond to the needs of individuals and communities.

☐ Identifies local/state needs and facilitates educational programs with measurable goals and objectives based on identified needs. (Program advisory groups and/or other means of providing external input are part of the process.)

☐ Incorporates evaluation processes into program plan that will adequately demonstrate the impact of the program.

Comments:

Goals:

Program Management:

☐ Delivers programs in an effective and efficient manner; organizes processes that result in quality program delivery.

☐ Demonstrates ability to involve, manage, facilitate and/or coordinate volunteer and/or paid staff for effective program delivery.

☐ Secures and manages alternative revenue sources to optimize program scope or size to obtain educational outcomes that are appropriate for program assignment.

Comments:

Goals:

Volunteer/Cooperator/Stakeholder Involvement:

☐ Identifies needs and opportunities for volunteer/cooperator/stakeholder service and partnership; recruits volunteers/cooperators to fill appropriate roles.

☐ Supports volunteers/cooperators/stakeholders through appropriate orientation, recognition, and evaluation.

Comments:

Goals:

2. Program Delivery

Teaching Skills

- ☐ Communicates effectively with individuals and groups based on evidence provided by teaching evaluations. (Minimum of three teaching evaluations required annually. Peer teaching reviews, approved by immediate supervisor and department head, are also encouraged.)

- ☐ Designs educational programs to support behavior change among participants, documented through impact assessment.

Comments:

Goals:

Educational Methodology and Communication Tools:

- ☐ Develops and/or adapts a variety of high quality communication tools to interest and inform target audiences.

- ☐ Uses educational methodologies appropriate for target audiences.

- ☐ Clearly documents the impact of educational activities on the economic well-being, environmental quality, and/or quality of life of target audience.

Comments:

Goals:

Scholarship and Creative Activity

Scholarship and creative activity are understood to be intellectual work whose significance is validated by peers and which is communicated. (See academic home's promotion and tenure guidelines for specific details.)

Levels of Scholarship:

☐ Interprets and integrates research-based knowledge from multiple sources, bringing new insights to issues. Validation by peers and communication are necessary steps to achieving scholarship.

☐ Participates in intellectual work or other creative activity (applied research, new teaching tools, competitive grants, innovative curriculum, evaluative processes, etc.), which is validated by peers and shared with other professionals.

☐ Initiates and/or provides leadership for intellectual work or other creative activity (applied research, new teaching tools, competitive grants, innovative curriculum, evaluative processes, etc.), which is validated by peers and shared with other professionals.

Comments:

Goals:

Documented Subject Matter Knowledge and Expertise:

☐ Possesses a recognized expertise in a narrow or broadly defined subject matter area of importance to current Extension programming.

☐ Possesses a recognized expertise that is utilized by OSU faculty as a resource to the discipline.

☐ Documents (through published articles, invited presentation, etc.) recognition of expertise beyond OSU and Oregon.

Comments:

Goals:

Interpersonal Skills

Local Faculty/Staff Relationships:

☐ Interacts effectively and appropriately with all individuals in the office.

☐ Makes contribution to office team efforts; valued as a team member.

☐ Provides leadership for building relationships and fosters open communications among staff and faculty.

Comments:

Goals:

Peer Relationships:

☐ Engenders trust and respect among peers.

☐ Cooperates across program areas as appropriate for position description.

☐ Demonstrates leadership in team situation.

Comments:

Goals:

Clientele Relationships:

☐ Relates effectively and appropriately with local clientele, volunteers, and collaborators.

☐ Engenders trust and respect among clientele.

☐ Recognized as leader among clientele and within community.

Comments:

Goals:

Community Relationships:

☐ Promotes a positive, professional image of Extension and Oregon State University.

☐ Develops strong positive community contacts for Extension and Oregon State University.

☐ Markets Extension as a part of the greater Oregon State University system and clearly communicates how Extension fits into the university's mission.

Comments:

Goals:

Instructions: For the next three categories, each performance dimension requires either a "yes" or "no" response. Evaluative comments should relate to the category.

Professional Development It is through life-long learning and growth that Extension faculty (check the appropriate responses): members meet the expectation to maintain program quality and relevance throughout their employment with OSU Extension.

☐ Yes ☐ No

• Develops a written professional development plan with appropriate venues for professional growth, including relevant sessions at annual conference, in-service training, course-work, academic meetings and other developmental experiences.

☐ Yes ☐ No

• Attends Extension Annual Conference and program related in-service training.

☐ Yes ☐ No

• Participates in professional meetings and other venues to learn from others.

Comments:

Goals:

Service Faculty service is essential to the University's success in achieving its central mission. Service is an expectation of all faculty members at Oregon State University.

University Service (check the appropriate response):

☐ Yes ☐ No

• Serves on Extension (at local and/or system levels), department, college and/or university committees and/or contributes to Extension/university relations in the community.

Service to the Profession (check the appropriate response):

☐ Yes ☐ No

• Serves as an active member of appropriate professional associations/societies at the state, regional and/or national level.

Service to the Public (check the appropriate response):

☐ Yes ☐ No

• Develops an appropriate balance of service to community that contributes to Extension/university relations, both within and outside assignment.

Comments:

Goals:

Organizational Accountability and Maintenance

Inclusiveness (check the appropriate response):

☐ Yes ☐ No

• Documents a balance of participant diversity within program assignment that is equal to or greater than the appropriate population statistics.

Reporting (check the appropriate response):

• Completes all requested reports in a comprehensive and timely manner. ☐ Yes ☐ No

Evaluation (check the appropriate response):

• Documents measurable program outcomes on at least one priority educational effort annually. ☐ Yes ☐ No

Comments:

Goals:

Overall Evaluation: (Must be completed for all faculty.) Considering the performance standards outlined in this form (including all the performance dimensions and indicators), the position description, and length of service, rank, and experience in the current assignment, check the definition that best describes this faculty member's overall performance during the evaluation period.

☐ **Extraordinary Performance:** Consistently and significantly exceeds performance standards throughout the review period. Performance is conspicuously marked by distinction, achieving excellence in all relevant performance dimensions. Individual exercises exceptional leadership in several areas of work responsibility.

☐ **Strong and Positive Performance:** Often exceeds performance standards, with outstanding contributions in Teaching, Facilitation, and Other Assignments; Scholarship and Creative Activity; and Interpersonal Skills. Performance is fully satisfactory.

☐ **Inconsistent Performance:** Meets performance standards in some areas, but not in others. Performance is not fully satisfactory. Improvement is needed in certain aspects of job responsibilities to meet expected performance standards.

☐ **Unsatisfactory Performance:** Consistently does not meet standards of performance. Performance of job responsibilities is unsatisfactory and unacceptable. (Significant improvement is needed to meet job responsibilities and to maintain employment.)

List faculty member's professional goals for coming year: (Summarize professional goals listed for each performance dimension here. These goals should be jointly agreed upon between the faculty member and the immediate supervisor, with progress made toward achieving the goals reviewed in next year's performance appraisal.)
Summary of evaluation by immediate supervisor:

Faculty member's comments: (optional)

Administrative Leadership (For Staff Chairs only.) This portion of the form is specific for evaluating staff chair responsibilities. There are overlapping performance expectations for staff chair and program assignments, such as the categories of Interpersonal Skills, Professional Development, and Service. Performance relating to both staff chair and program responsibilities will be evaluated on the pervious pages of this form.

Instructions: Objective, constructive feedback is essential to the development of faculty and team productivity and performance. Using this instrument as a guide, discuss each performance dimension, including the performance indicators listed, the employee's demonstrated skill and ability in that area, and any suggestions for development. Check the performance indicators the faculty member has achieved in this evaluation period. (To make a check, point cursor on lower, right-hand corner of box and left click mouse.) Comments should relate to the performance indicators, identifying specific strengths and areas for improvement. Comments are to be based on the current job description and the faculty member's experience and length of service in the assignment. Professional goals listed for each performance dimension should be jointly agreed upon between the faculty member and the immediate supervisor, with progress made toward achieving the goals reviewed in next year's performance review.

Program Leadership:

☐ Provides vision, direction, and priorities for overall Extension program in assigned county(ies).

☐ Establishes effective organizational structure within the work unit; organizes processes for effectiveness and efficiency.

☐ Makes effective decisions that are respected by others and developed in a manner that invites and incorporates appropriate input from others.

Comments:

Goals:

Stakeholder Relationships:

☐ Monitors stakeholder satisfaction; addresses stakeholder concerns; is accountable to stakeholders.

☐ Manages a broadly representative citizens advisory group.

☐ Communicates with county, state, and federal decision makers with a consistent, positive message.

☐ Markets OSU Extension to be the premier provider of nonformal education that meets the needs of Oregonians and to create and maintain a positive perception among community stakeholders.

Comments:

Goals:

Human Resource Management:

☐ Recruits, motivates, and retains diverse faculty and staff.

☐ Maintains high standards and provides criterion-based performance evaluation, encouraging, recognizing and rewarding improvement and achievement.

☐ Manages performance problems promptly and appropriately.

☐ Maintains a well functioning, service-oriented office.

☐ Provides guidance, as appropriate, to volunteer management efforts of faculty and staff.

Comments:

Goals:

Financial Management:

☐ Develops and secures budget to provide adequate support and services.

☐ Manages budget to accomplish program objectives with the available resources.

☐ Demonstrates ability to analyze, judge and deliver sound fiscal decisions in concert with others.

☐ Reports to decision makers in a fiscally responsible and accountable manner.

Comments:

Goals:

Organizational Support:

☐ Demonstrates knowledge of organizational policies and procedures.

☐ Represents faculty, staff and county to administration, and represents administration to faculty, staff and county.

☐ Works as a team member with administration in support of organizational decisions and policies.

Comments:

Goals:

APPENDIX G RESPONSE RATE BY STATE

State	Number of potential respondents	Number of actual respondents	Percent of total respondents	Percent response rate per state
Alaska	4	2	1	50
Alabama	5	3	1	60
Arizona	11	6	3	55
Arkansas	6	0	0	0
California	8	5	2	63
Colorado	72	36	16	50
Connecticut	3	1	0	33
Delaware	2	2	1	100
Florida	12	6	3	50
Georgia	10	2	1	20
Hawaii	7	2	1	29
Illinois	1	0	0	0
Idaho	6	3	1	50
Iowa	4	0	0	0
Indiana	13	3	1	23
Kansas	2	1	0	50
Kentucky	10	4	2	40
Louisiana	9	6	3	67
Maine	1	0	0	0
Maryland	1	0	0	0
Massachusetts	14	2	1	14
Michigan	40	19	9	48
Missouri	2	0	0	0
Mississippi	14	6	3	43
Minnesota	16	8	4	50
Montana	2	2	1	100
Nevada	6	3	1	50
Nebraska	11	1	0	9
New Hampshire	12	5	2	42
New Jersey	3	2	1	67
New Mexico	1	0	0	0
New York	54	11	5	20
North Carolina	12	7	3	58
North Dakota	8	2	1	25
Ohio	10	3	1	30
Oklahoma	4	2	1	50
Oregon	23	10	5	43
Pennsylvania	20	13	6	65
Rhode Island	3	1	0	33
South Carolina	9	7	3	78
South Dakota	1	0	0	0
Tennessee	5	3	1	60
Texas	12	4	2	33
Utah	6	5	2	83
Vermont	8	3	1	38
West Virginia	3	1	0	33
Virginia	15	6	3	40
Washington	13	5	2	39

State	Number of potential respondents	Number of actual respondents	Percent of total respondents	Percent response rate per state
Wisconsin	8	3	1	38
Wyoming	1	0	0	0
Other	-	3	1	-
Total	**523**	**219**	**100**	**-**

LITERATURE CITED

Afrea (2004). African Evaluation Association [on line]. Available from http://www.afrea.org. Cited on July 12, 2004.

Ahearn, M., J. Yee, and J. Bottum. (2003). Regional Trends in Extension System Resources, United States Department of Agriculture. Agriculture Information Bulletin No. 781 [on line]. Available from http://www.ers.usda.gov. Cited on July 12, 2004.

Andersen, A. E. (1997). Forestry Extension Infrastructure for the Non-Industrial Private Forestry Sector in the Nordic Countries. Unpublished report. Vancouver, B.C.

AREERA (2004). Agriculture Research, Extension and Education Reform Act [on line]. Available from http:// www.reeusda.gov/part/areera. Cited on July 12, 2004.

Ary, D., L. C. Jacobs, and A. Razavieh (1996). Introduction to Research in Education, Harcourt Brace College Publishers.

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

Bandura, A. (1977). Social Learning Theory. Englewood Cliffs, New Jersey.

Barnette, J. J. and J. R. Sanders (2003). The Mainstreaming of Evaluation, Jossey-Bass, San Francisco.

Beeman, C. E., J. G. Cheek, M.B. McGhee, and E. M. Grygotis (1979). Professional Competencies Needed by Extension Agents in the Florida Cooperative Extension Service: A report of research. Department of Agriculture and Extension Education, University of Florida.

Begus, J. (2004). Personal Communications. May 20, 2004.

Bennett, C. F. (1976). Analyzing Impacts of Extension Programs. Washington D.C., U.S. Department of Agriculture Extension Service, No. ESC 575.

Bennett, C. and K. Rockwell (1995). Targeting Outcomes of Programs (TOP): An integrated approach to planning and evaluation. Unpublished manuscript, Lincoln, NE: University of Nebraska.

Bennett, C. and K. Rockwell (2004) Targeting Outcomes of Programs [on line]. Available from http://citnews.unl.edu/TOP/english/synopsis . Cited on Sept. 17, 2004.

Blackburn, R. T. and J. H. Lawrence (1995). Faculty at Work: Motivation, Expectation, Satisfaction. John Hopkins University Press, Baltimore.

Boon, T. E. (2003). Attitudes of Danish Forest Owners, Danish Forest and Landscape Research Institute, Hoersholm, Denmark.

Brody, N. (1983). Human Motivation: Commentary on Goal-Directed Action. Academic Press, New York.

BTAA (2000). Budget Transparency and Accountability Act of British Columbia, Province of British Columbia.

Bush, C., R. Mullis, and A. Mullis (1995). "Evaluation: An Afterthought or an Integral Part of Program Development." Journal of Extension **33**(2) [on line]. Available from http://www.joe.org/joe/1995april/a4.html. Cited on July 12, 2004.

CDC (Centers for Disease Control and Prevention). (2004) Framework for Program Evaluation in Public Health. MMWR 1999;48, No. RR-11.

Compton, D., M. Baizerman, S. Stockdill (2002). The Art, Craft, and Science of Evaluation Capacity Building. Jossey Bass, San Francisco.

Cooper, A. and D. Graham (2001). "Competencies Needed to Be Successful County Agents and County Supervisors." Journal of Extension **39**(1) [on line]. Available from http://www.joe.org/joe/2001february/rb3.html. Cited on July 12, 2004.

Decker, D. (1990). "Organizational philosophy for program evaluation." Journal of Extension **28**(2) [on line]. Available from http://www.joe.org/joe/1990summer/f1.html. Cited on July 12, 2004.

Dillman, D. A. and P. Salant (1994). How To Conduct Your Own Survey. John Wiley and Sons, Inc., New York.

Douglah, M., H. Boyd, D. Gundermann (2003). Nurturing the Development of An Evaluation Culture in Public Education Agencies. Paper presented at the American Evaluation Association Annual Conference, November 5-8, 2003, Reno, NV.

Doyle, R. and D. Lemaire, Eds. (1999). Building Effective Evaluation Capacity: Lessons From Practice. Transaction Publishers, New Jersey.

Duigan, P. (2003). Mainstreaming Evaluation or Building Evaluation Capacity? Three Key Elements. The Mainstreaming of Evaluation. *Edited by* J. J. Barnette and J. Sanders. San Francisco, Jossey Bass.

ECOP (a) (2000). The Extension System: A Vision for the 21st Century, Extension Committee on Organization and Policy, Cooperative State Research and Extension Education Service.

ECOP (b) (2000). Guidelines for Professional Development in Extension, Extension Committee on Organization and Policy, Cooperative State Research and Extension Education Service.

Elder, G. H., Jr. (1985). Life Course Dynamics: Trajectories and Transitions, 1968-1980. Cornell University Press, Ithaca.

FAO (1997). Food and Agriculture Organization of the United Nations. Issues and Opportunities in the Evolution of Private Forestry and Forestry Extension in Several Countries in Transition in Central and Eastern Europe. Rome [on line]. Available from http://www.fao.org/docrep/w7170E/w7170E00.htm,

Featherman, D. L. (1983). Life Span research in Social Science Research. Life Span Development and Behaviour. *Edited by* P. B. Baltes and J. O. G. Brim. Academic Press, New York.

Ferlie, E., L. Ashburner, L. Fitzgerald, and A. Pettigrew (1996). The New Public Management in Action. Oxford University Press, Oxford.

Ferlie, E. and J. Hartley (2003). "Changing Public Service Organizations: Current Perspectives and Future Prospects." British Journal of Management **14:** (S1-14).

Fetterman, D. M. (2000). Foundations of Empowerment Evaluation. Sage Publications, Thousand Oaks.

Field, A. (2000). Discovering Statistics Using SPSS for Windows. Sage Publications, London.

Forest, L. B., B. E. Rossing and C.C. Coggins. (1989). Accountability: The Challenge and Response. Foundations and Changing Practices in Extension. *Edited by* D. B. Blackburn, University of Guelph, Guelph.

FORREX (2004). Forest Research Extension Service [on line]. Available from http://www.forrex.org. Cited on July 12, 2004.

FWAG (2004). Farming and Wildlife Advisory Group Home Page [on line]. Available from http://www.fwag.org.uk. Cited on July 12, 1004.

Geering, A. D. (1980). The Current State of Research on Motivation. Adelaide College of the Arts and Education, Australia. U.S. Dept. of Education, Educational Resources Information Center (ERIC) No. ED196911.

GPRA (1993). Government Performance and Results Act [on line]. Available from http://www.whitehouse.gov/omb/mgmt-gpra/gplaw2m.html. Cited on July 12, 2004.

Guba, E. G. and Y. S. Lincoln (1989). The Coming of Age of Evaluation. *In* Fourth Generation Evaluation. Sage Publications, Newbury Park.

GCT (2004). Game Conservancy Trust [on line]. Available from http://www.gct.org.uk/. Cited on July 14, 2004.

Hair, J. F., R. E. Anderson, R. L. Tatham, and W.C. Black (1998). Multivariate Data Analysis. Prentice Hall, Upper Saddle River.

Hamilton, R. A. and L. E. Biles (1998). Forestry Extension in the United States. In third IUFRO Extension Working Group Symposium, Blacksburg, Virginia.

Henry P. Sims, J. and P. Lorenzi (1992). The New Leadership Paradigm: Social Learning and Cognition in Organizations. Sage Publications, Newbury Park.

Herzberg, F. (1991). One more Time: How do you Motivate Employees? Harvard Business Review, Boston. Paperback No 90010.

Hoberg, G. (2002). Finding the Right Balance: Designing Policies for Sustainable Forestry in the New Era. Vancouver, B.C., UBC Faculty of Forestry Jubilee Lecture Series.

Horton, D., A. Alexaki, S. Bennett-Lartey, K.N Brice, D. Campilan, F. Carden, J. De Souza Silva, L.T. Doung, I. Khadar, A. Maestrey Boza, I. Kayes Muniruzzaman, J. Perez, M. Somarriba Chang, R. Vernooy, and J. Watts. (2003). Evaluating Capacity Development: Experiences from Research and Development Organizations around the World. The Netherlands: International Service for National Agricultural Research (ISNAR), International Development Research Centre (IDRC), and ACP-EU Technical Centre for Agricultural and Rural Cooperation.

Hoy, W. and C. Miskel (1978). Educational Administration: Theory, Research, and Practice. Random House, New York.

International Development Research Council (2004) In Conversation with Michael Quinn Patton, http://web.idrc.ca/en/ev-30442-201-1-DO_TOPIC.html (2004). Cited on July 12, 2004.

IAPE (2004). Israeli Association for Program Evaluation [on line]. Available from http://ayelet.netfirms.com. Cited on July 12, 2004.

IDEAS (2004). International Development Evaluation Association [on line]. Available from http://www.ideas-int.org/Index.aspx. Cited on July 12, 2004.

IEPEC (2004). International Energy Program Evaluation Conference [on line]. Available from www.iepec.org/evalorgs. Cited on July 12, 2004.

IPEN (2004). Russian International Project Evaluation Society [on line]. Available from http://ipen21.org/ipen.html. Cited on July 12, 2004.

IOCE (2004) The International Organisation for Cooperation in Evaluation [on line]. Available from http://home.wmis.net/~russon/ioce. Cited on July 12, 2004.

Jha, L. R. (2001). Using Appreciative Inquiry to Test the Application of Outcome Engineering in Extension Programs. PhD dissertation. University of Nebraska- Lincoln. [on line]. Available from http://aglec.unl.edu/aglecrevise/jhaabstract.htm. Cited on July 12, 2004.

Katzell, R. A. and D. E. Thompson (1990). "Work Motivation: Theory and Practice." American Psychologist **45**(2): 144-153.

Kellogg Commission (2001). Draft Report of the "Extension Vision for the 21st Century Committee. Battle Creek, Michigan.

Kellogg Foundation (February 1999). Returning to Our Roots: The Engaged Institution, National Association of State Universities and Land-Grant Colleges. [on line]. Available at http://www.osu.edu/osu/newsrel/Archive/99-03-01_OSU_Engagement_Highlighted_by_Kellogg_Report.html. Cited on July 12, 2004.

King, J. (2002). Building the Evaluation Capacity of a School District. The Art, Craft, and Science of Evaluation Capacity Building. *Edited by* D. Compton, M. Baizerman and S. H. Stockdill. Jossey-Bass, San Francisco.

Kohn, A. (1993). "Why Incentive Plans Cannot Work." Harvard Business Review, September-October.

Lawton, A. and A. Rose (1994). Organisation and Management in the Public Sector. Pitman Publishing, London.

Levine, J. (2001). Core Competency requirements. Michigan State University Extension Service, East Lansing, Michigan.

Lindner, J. R. (1998). "Understanding Employee Motivation." Journal of Extension **36**(3). [on line]. Available from http://www.joe.org/joe/1990summer/f1.html. Cited July 12, 2004.

Locke, E. A. and G. P. Latham (1994). Goal Setting Theory. Motivation: Theory and Research. *Edited by* J. Harold F. O'Neil and M. Drillings. Lawrence Erlbaum Associates, Publishers, Hillsdale.

Love, A. (1991). Internal Evaluation: Building Organizations from Within. Sage Publications, Newbury Park.

Luthans, F. (1995). Organizational Behavior. New York, McGraw Hill.

McDavid, J. (2001). Program Evaluation in British Columbia in a Time of Transition: 1995-2000. Canadian Journal of Program Evaluation, Special Issue. (p. 3-28).

Mackay, K. (2002). The World Bank's Evaluation Capacity Building Experience. *In* The Art, Craft, and Science of Evaluation Capacity Building. *Edited by* D. Compton, M. Baizerman and S. H. Stockdill. Jossey-Bass, San Francisco.

Maeher, M. L. and L. A. Braskamp (1986). The Motivation Factor: A Theory of Personal Investment. University of Chicago Press, Chicago.

Maddock, R. C. and R. L. Fulton (1998). Motivation, Emotions and Leadership. Quorum Books, Westport.

Milstein, B., T. J. Chapel,S. Wetterhall, and D. Cotton. (2002). Building Capacity for Program Evaluation at the Centers for Disease Control and Prevention. *In* The Art, Craft, and Science of Evaluation Capacity Building. *Edited by* D. Compton, M. Baizerman and S. H. Stockdill.

Milstein, B. and D. Cotton (2000). Defining Concepts for the Presidential Strand on Building Evaluation Capacity. American Evaluation Association. [on line]. Available from http://www.eval.org/eval2000/public/pressstrand.pdf. Cited on July 12, 2004.

Ministry of Forests (2004a). British Columbia Ministry of Forests Home Page [on line]. Available from http://www.for.gov.bc.ca

Ministry of Forests (2004b). Forest and Range Practices Act [on line]. Available from www.for.gov.bc.ca/code/ 2004. Cited on July 12, 2004.

Mohamed, A. R. A. (1998). County Agent's Assessment of the Ohio State University Extension Support and Performance Evaluation System. PhD dissertation. The Ohio State University.

ODI (2004). Neuchatel Initiative, Overseas Development Institute [on line]. Available from http://www.odi.org.uk/rpeg/Ni_index.html. Cited on July 12, 2004.

O'Neil, H. F. and M. Drillings (1994). Motivation: Theory and Research. Lawrence Erlbaum Associates, Hillsdale.

OSU Extension (2001). Public Issues Education, Oregon State University Extension Service [on line]. Available from http://oregonstate.edu/dept/pie/. Cited on July 13, 2004.

Osborne, J. and E. Waters (2002). Four assumptions of multiple regression that researchers should always test [on line]. Available from http://PAREonline.net.getvn.asp?v=8&n=2. Cited on July 12, 2004.

Pardee, R. L. (1990). Motivation Theories of Maslow, Herzberg, McGregor, and McClelland: A Literature Review of Selected Theories Dealing with Job Satisfaction and Motivation, U.S. Dept. of Education, Educational Resources Information Center (ERIC).

Patton, M. Q. (1997). Utilization-Focused Evaluation. Sage Publications, Thousand Oaks.

Patton, M. Q. (1987). "The Extension Organization of the Future." Journal of Extension **25**(1) [on line]. Available from http://www.joe.org/joe/1987spring/fut1.html. Cited on July 13, 2004.

Pavlov, I.P. (1960). Conditioned Reflexes: An investigation of the Physiological Activity of the Cerebral Cortex. Dover Publications, New York.

Pinder, C. (1984). Work Motivation: Theory, Issues, and Application, Harper Collins.

Preskill, H. and D. Russ-Eft (2004). Building Evaluation Capacity: 72 Activities for Teaching and Training. Sage Publications, Thousand Oaks.

Radhakrishna, R. and M. Martin (1999). "Program Evaluation and Accountability Training Needs of Extension Agents." Journal of Extension **27**(3) [on line]. Available from http://www.joe.org/joe/1999june/rb1.html. Cited on July 12, 2004.

Reid, R. and P. Stephen (2002). The Australian Master TreeGrower Program1996-200, Development, delivery and impact of a national outreach and education program. A report for the RIRDC/L&W Australia/FWPRDC Joint Venture Agroforestry Program. Melbourne, Australia.

Ritsos, P. B. and L. E. Miller (1985). Professional Competencies Needed by Extension Employees in Urban Counties of Ohio. Summary of Research 43. Columbus, Ohio, Ohio State University, Dept. of Agricultural Education.

RREA (1978). Renewable Resources Extension Act of 1978 [on line]. Available from http://www.reeusda.gov/1700/legis/renresex.htm. Cited on July 12, 2004.

Sanders, J. R., Ed. (1994). The Program Evaluation Standards, 2nd Edition. Joint Committee on Educational Evaluation. Sage Publications, Thousand Oaks.

Sanderson, D. R. (1988). Understanding Cooperative Extension: Our Origins, Our Opportunities. Working With Our Publics, Module 1. Raleigh, North Carolina, North Carolina State University Extension Service/Dept. of Adult and Community College Education.

Schedler, K. (1995) Ansaetze einer wirkungsorientierten Verwaltungsfuehrung. Von der Idee des New Public Managements (NPM) zum konkreten Gestaltungsmodell. Fallbeispiel Schweiz. Verlag Paul Haupt, Bern.

Scheirer, M. A. (1994). Designing and Using Process Evaluation. Handbook of Practical Evaluation, Jossey Bass, San Francisco.

Schonlau, M., J. Ronald D. Fricker, et al. (2002). Conducting Research Surveys via E-mail and the Web. Santa Monica, Rand.

Scriven, M. (1991). Evaluation Thesaurus. Sage Publications, Thousand Oaks.

Senge, P. (1990). The Fifth Discipline. Doubleday/Currency, New York.

Sims, H. J. and P. Lorenzi (1992). The New Leadership Paradigm: Social Learning and Cognition in Organizations. Sage Publications, Newbury Park.

Skinner, B. F. (1953). Science and Human Behaviour. Free Press, New York.

Sonnichsen, R. C. (2000). High Impact Internal Evaluation. Sage Publications, Thousand Oaks.

Spector, P. E. (1992). Summated Rating Scale Construction. Sage Publications, Newbury Park, California.

Summerhill, W. R. and C. L. Taylor (1986). What We Have Learned About Extension Faculty Adoption of Evaluation as a Basis for Program Accountability. Annual meeting of American Evaluation Association, Kansas City, Missouri.

Sutphin, C. and J. Hillison (1999). "History of Extension Work in Virginia Prior to Smith-Lever." Journal of Extension **37**(6) [on line]. Available from http://www.joe.org/joe/1999december/comm2.html. Cited on July 12, 2004.

Suvedi, M., C. den Biggelaar, and S. Morford (2003). "Conceptual Framework for Evaluating Sustainable Agriculture." Journal of Crop Production 9(1/2): 433-454.

Suvedi, M. and S. Morford (2003). Conducting Program and Project Evaluations: A Primer for Natural Resource Program Managers in British Columbia. Kamloops, B.C., FORREX (Forest Research Extension Partnership) [on line]. Available at http://www.forrex.org/pubs. Cited on July 12, 2004.

Taylor-Powell, E. (2002). Competence in Extension Education Evaluation- What is it? What Capacity Does it Entail? Extension Evaluation Topical Interest Group: Hear It From the Board.

Torres, R. and H. Preskill (2001). "Evaluation and Organizational Learning: Past, Present, and Future." American Journal of Evaluation **22**(3): 387-395.

Toulemonde, J. (1999). Incentives, Constraints and Culture-building as Instruments for the Development of Evaluation Demand. Building Effective Evaluation Capacity: Lessons From Practice. Transaction Publishers, New Jersey.

United Way of Greater Toronto. (2003). Capacity Building for Program Effectiveness. Paper presented at the Canadian Evaluation Society conference, Vancouver, British Columbia.

U.S. Department of Agriculture (2001). The Cooperative State Research, Education and Extension Service of the USDA Home Page, U.S. Department of Agriculture [on line]. Available from http://www.ueeusda.gov/1700/csrees_main.htm. Cited on July 12, 2004.

Wandersman, A., P. Flaspohler, A. Ace, L. Ford, P. Imm, M. Chinman, J. Sheldon, A.B. Andrews, C. Crusto, J. Kaufman (2003). PIE a la mode: Mainstreaming Evaluation and Accountability in Each Program in Every County of a Statewide School Readiness Initiative. The Mainstreaming of Evaluation. Edited by J. J. Barnette and J. R. Sanders. Jossey-Bass, San Francisco.

Warner, P. D. and J. A. Christenson (1984). The Cooperative Extension Service: A National Assessment. Westview Press, Boulder and London,.

Williams, D. D. and M. L. Hawkes (2003). Issues and Practices Related to Mainstreaming Evaluation: Where Do We Flow from Here? The Mainstreaming of Evaluation. J. J. Barnette and J. Sanders. Jossey Bass, San Francisco.

Wiseman, F. (2003). "On the Reporting of Response Rates in Extension Research." Journal of Extension **41**(3) [on line]. Available from http://www.joe.org/joe/2003june/comm1.shtml. Cited on July 12, 2004.

World Bank (2002). Annual Report on Evaluation Capacity Development. Washington D.C., Operations Evaluation Department.