The development and evaluation of:
A Guidebook for Evaluating Residential Outdoor
Education Programs (R.O.P.E.)

A Thesis

Presented in Partial Fulfillment of the Requirements for
the degree of Master of Science in the Graduate School
of The Ohio State University
by
Heather Lee Martin, B.S.

The Ohio State University
1994

Master's Examination Committee:
Dr. Gary W. Mullins, Chair
Dr. Emmalou Norland
Dr. Robert E. Roth

Approved by
Gary W. Mullins
Advisor
School of Natural Resources
DEDICATION

To my parents who taught me to never say "I can't."
ACKNOWLEDGMENTS

I would like to express my gratitude to all who have encouraged me to grow through this learning process. I was blessed with a special advisor, Dr. Gary Mullins. Much appreciation goes to you for your tremendous support and concern for students. I also extend my gratitude to the other members of my committee, Dr. Emmalou Norland and Dr. Robert Roth.

Sincere thanks go to the panelists who offered their time and provided insightful comments on the guidebook. This project was dependent on your contributions. Gratitude is also extended to The Ohio State University, the Ohio Agriculture Research and Development Center, and the faculty of the School of Natural Resources for both academic and physical support.

I wish to thank my family who made this possible. To my mother I owe great thanks for teaching me how to focus and to believe in myself. To my father, thanks for your steady support. To my brothers who challenged me to always do my best, I owe gratitude. To Laddin, thanks for your endless support, love, and understanding. Additional thanks go to my friends who provided much encouragement throughout the process.
VITA

October 21, 1970 ........................................ Born - Gloucester, Virginia

1989........................................................... Cabin Counselor, Makamie Woods, Barhamsville, Virginia

1990........................................................... State Park Interpreter, Westmoreland State Park, Montross, Virginia

1991........................................................... Research Assistant, Biology Department, University of Alabama, Tuscaloosa, Alabama

1992........................................................... B.S., Biology, Mary Washington College, Fredericksburg, Virginia


1992-1993.................................................. Graduate Fellow, School of Natural Resources, The Ohio State University, Columbus, Ohio

1993........................................................... Group Counselor, Pine Hill Adventures, Priceville, Alabama

1994........................................................... Graduate Research Associate, School of Natural Resources, The Ohio State University, Columbus, Ohio

1994........................................................... Environmental Educator, Camp McDowell Environmental Education Center, Nauvoo, AL

1994........................................................... Private Consultant, Martin & Associates

FIELDS OF STUDY

Major Field: Natural Resources

Studies in:

Environmental Communication, Education and Interpretation Evaluation
# TABLE OF CONTENTS

DEDICATION .................................................................................................................. .ii  
ACKNOWLEDGMENTS ................................................................................................... iii  
VITA ................................................................................................................................ .iv  
TABLE OF CONTENTS .................................................................................................... v  
LIST OF TABLES............................................................................................................ vii  
LIST OF FIGURES.......................................................................................................... viii  

CHAPTER  

I. INTRODUCTION........................................................................................................... 9  
  Purpose, Objective, and Research Questions ......................................................... 11  
  Definition of Terms ................................................................................................. 12  
  Limitations of the Study ......................................................................................... 13  
  Basic Assumptions ................................................................................................. 14  

II. REVIEW OF THE LITERATURE ............................................................................ 15  
  Camping Literature ............................................................................................... 14  
  Evaluation of Resident Outdoor Education Programs ............................................ 22  
  Evaluation Literature Summary ............................................................................ 43  

III. PROCEDURES......................................................................................................... 46  
  Research Design ..................................................................................................... 46  
  Subject Selection .................................................................................................... 51  
  Outcome Measures ................................................................................................. 52  
  Instrumentation ....................................................................................................... 55  
  Instrument Development ......................................................................................... 57  
  Conditions ................................................................................................................ 61  
  Data Analysis ........................................................................................................... 62  

IV. RESULTS ................................................................................................................ 63  
  Findings From the Written Questionnaire ............................................................... 64  
  Findings From Comments on the Guidebook ........................................................... 85  
  Summary of Changes Made to the Draft ................................................................. 94
V. CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS

Conclusions about the R.O.P.E. model ...................................................... 96
Discussion of the Research Process ............................................................ 97
Recommendations for further modifications of the guidebook ................. 103
Suggestions for future research and developments ............................... 104

LIST OF REFERENCES ................................................................................. 108

APPENDICES

A Panel of experts participating in the pilot and field tests ...................... 115
B Letter to the field test participants ....................................................... 117
C Request letter to the academic panelists .............................................. 119
D Cover letter to the academic panelists ............................................... 121
E Questionnaire distributed to the academic panelists ......................... 123
F Cover letter to the field expert panelists .............................................. 130
G Questionnaire distributed to the field expert panelists ...................... 132
H Follow-up postcard ............................................................................. 139
I Panelists .................................................................................................. 141
J Time line ................................................................................................ 143
K A Guidebook for Evaluating Residential Outdoor Education Programs
(R.O.P.E.) .................................................................................................. 145
LIST OF TABLES

TABLE

1. Field Experts' Responses to the Presentation of the Guidebook ................. 69
2. Field Experts' Responses to the Content of the Guidebook ......................... 71
3. Field Experts' Responses to the Performance of the Guidebook .................. 73
4. Correlation Coefficients for Data from the Panel of Field Experts ............... 75
5. Academic Experts' Responses to the Presentation of the Guidebook ............. 77
6. Responses of the Academic Experts to the Content of the Guidebook ............ 79
7. Academic Experts' Responses to the Performance of the Guidebook ............. 81
8. Correlation Coefficients for Data from the Panel of Field Experts ............... 83
9. Field Experts' Comments Written on the Guidebook .................................. 86
10. Academic Experts' Comments on the Content ........................................ 89
11. Academic Experts' Comments on the Presentation .................................... 90
LIST OF FIGURES

FIGURE

1. Bennett's Hierarchy of Evidence For Program Evaluation.......................... 35
CHAPTER I
INTRODUCTION

Residential outdoor education programs, which utilize out-of-doors settings to teach students, have evolved around the tenant that these programs provide unique benefits to the attendees. Many specific types of programs in the United States endorse residential outdoor education; these include organization sponsored summer camps, private independent camps, outdoor adventure camps, and, more recently, environmental education camps. Positive effects that have been credited to these programs include: increased self-esteem; improved social skills; positive changes in attitudes towards the environment, school, and peers; cognitive gains; the development of new life skills; and increased environmentally appropriate behaviors. Research addressing the attainment of these effects or the specific goals of the camps, however, has been inconclusive (Hammerman, 1980; Roth, 1976). The programs' actual effects on the campers, which components of the programs promote the goals, and how program modifications may increase desired results remain in question.

For many years, evaluations of educational programs have been developed to aid program improvement (Cronbach, 1972; Geis, 1987; Patton, 1990). In the recent literature, an interest in the improvement of these camps through program evaluation has been voiced (Baker, 1983; Stone & Woods, 1981; Hollenhorst & Ewert, 1985; Henderson, 1992). Planned program evaluations in the field of outdoor education may
be used to gain a better understanding of the program (Ewert, 1987), to provide accountability (Henderson, 1987), and to aid in program and staff improvement (Chenery & Hammerman, 1985a).

Camp administrators and personnel are also concerned about the need for evaluation and research on camping. In a survey delivered to attendees of the 1983 American Camping Association (ACA) convention (primarily camping professionals), eighty-eight percent of the respondents indicated that the ACA should conduct research in the field of organized camping and seventy-five percent felt that research results should be reported in their *Camping Magazine* (Henderson & Bialeschki, 1984). In contrast to the high level of concern for research and evaluation, "only a few outdoor centers seemed to be actively engaged in evaluation" (Hammerman & Chenery, 1985b, p. 37). Thus, the concern for evaluation is present; yet, little progress is being made.

Through consultation of camp administrators, Henderson and Chenery (1987) found that, "Lack of time, hesitancy to intrude upon the camp experience, and questions about adequate measurements seem to create barriers to camp research."

These authors indicate that a more fortuitous barrier may be the lack of training in research and evaluation. In a recent work in conjunction with another researcher, Bialeschki, Henderson further explains the barrier created by lack of knowledge by stating:

> We suspect the major reasons evaluations are not systematically conducted in camps are because camp personnel do not know how to set up an effective evaluation process, how to analyze the data, and/or how to interpret the data in useful ways to assist in accountability and decision-making. (1993, p. 32)
One method to overcome these barriers is to provide camp administrators with a guide for evaluation techniques. As Hammerman and Chenery (1985b, p. 37) state:

There may be a need for a model to guide the development of a systematic evaluation and research process that can be tailored to individual outdoor education programs' situations and fit within the constraints of time and budget.

Residential outdoor education programs (ROEP, as distinguished from the guidebook, R.O.P.E.) could benefit from carefully planned program evaluations. Perhaps traditional educational evaluation designs could be used to assess the worth of these programs. However, residential programs encompass many educational opportunities not encountered in formal education; therefore, traditional educational evaluation plans may not be adequate for assessing the effects of these programs (Baker, 1983; Hammerman, 1981). A heuristic model for the evaluation of ROEP could assist administrators in the development of such evaluations. Currently, no systematic method for developing an appropriate evaluation strategy exists for residential outdoor education programs (ROEP). In conclusion, there is a need for the development and assessment of a systems model for the evaluation of ROEP.

**Purpose, Objective, and Research Questions**

The purpose of this study was to develop a decision-making model, supported by background information, which can aid ROEP administrators in the development of an evaluation plan for their particular program. The specific objective was:

To develop, evaluate, and refine a systems model for utilization of alternative evaluation techniques for residential outdoor education programs. Such a systems model will have perceived utility by the user, will be user friendly, and will capture the pragmatics of daily camp evaluation needs, yet it will be firmly grounded in relevant theory and sound research management practices.
The research questions for the study were:

1) What are the academic and field expert panelists' perceptions of the presentation, content (quality and quantity), and performance of the R.O.P.E. guidebook?

2) What is the relationship between the panelists' perceptions of the quality and the performance of the model and their perceptions of (1) overall presentation, (2) individual elements of the presentation, (3) overall performance, and (4) individual elements of the performance?

The primary outcome of this study was a peer-reviewed evaluation guidebook to be used by ROEP in the development of individually tailored evaluation plans. This guidebook was designed to help overcome the previously acknowledged barriers to research and evaluation by providing an understandable, informative, and efficient framework for program evaluators. If the decision-tree model provides adequate flexibility, it may be adaptable to other educational settings (i.e., adventure education, interpretive programs, day-use outdoor education programs, etc.), and it may be used by teachers, principles, and other educational administrators. In addition to the creation of the model, this study demonstrated a methodology for the development of a tool (the guidebook) to be used in outdoor or environmental education.

**Definition of Terms**

**Program Administrator:** the person(s) most directly in charge of the daily administration of the program and facilities.

**Environmental Management Education:** an educational process that encourages: awareness of environmental problems and solution alternatives, knowledge about both biophysical and sociocultural environments, and motivation to act in a responsible manner (Roth, 1969).
Evaluation: "a systematic investigation of the worth or merit of some object" (Joint Committee on the Standards for Educational Evaluation [JCSEE], 1981, p.12).

Model: a heuristic device intended to exemplify an ideal case.

Opinion: the verbal statement of an attitude, "the sum total of a man's inclinations and feelings, prejudice or bias, preconceived notions, ideas, fears, threats, and convictions about a specified topic" (Thurstone, 1970, p. 128).

Outdoor Education: "education in, for and about the outdoors" (Donaldson & Donaldson, 1958, p. 63).

Program Evaluation: "evaluations that assess educational activities which provide services on a continuing basis and often involve curricular offerings" (JCSEE, 1981, p. 12).

Residential Program: program at which a majority of the attendees remain on the site for at least three days.

Utility: the degree to which an item is useful.

Limitations of the Study

It must be recognized that the purpose of this study was not to create and assess a theoretical model but a heuristic one. This study was not intended to produce results that are generalizable to all resident outdoor education programs, rather it was intended to gain a better understanding of the general applicability of a decision-tree type evaluation guidebook to programs that may make use of it. The study used the respondents' written perceptions of the guidebook. This limits the results which indicate only the perceived utility as opposed to the actual utility.
Additionally, the publication process for the guidebook entailing, page layout, graphics, and design aspects, is beyond the scope of this thesis project.

**Basic Assumptions**

A fundamental assumption in this study will be that the panelists will provide honest feedback on the guidebook. The population from which the panelists will be selected are generally known for their willingness to help. Persons who are in the field of resident outdoor education typically spend their professional lives striving to improve these programs "for the good of the campers;" therefore, the assumption that they will freely contribute constructive criticisms is warranted. Additionally, it will be assumed that the instruments developed for this study are accurate measures of the respondents attitudes about the utility of the guidebook.
CHAPTER II
REVIEW OF THE LITERATURE

Camping Literature

A Definition: Organized resident outdoor education in the United States of America began in the 1860's with Frederick Gunn's organized school tent camping experience for boys (Markovits, 1977). Since that time, resident outdoor education programs have become a heterogeneous group. The diversity of foci of the programs can be understood through a review of some of the many titles and descriptors that have been applied to these programs. These include: school camps, resident outdoor schools, summer camps, therapeutic camps, adventure vacations, church camps, organization sponsored camps, environmental education camps, wilderness education camps, etc.

Literature on outdoor education offers numerous definitions and much debate as to the propriety of any single definition. A typical definition can be found in the *Dictionary of Concepts in Recreation and Leisure Studies*: "The use of the outdoors as a setting for the teaching of skills, affective lessons, and cognitive lessons that include, but are not limited to, those directly associated with the natural ENVIRONMENT" (Smith, 1990, p. 226).
A classic definition, stating that outdoor education is "education in, about and for the outdoors" (Donaldson & Donaldson, 1958, p. 63), has been criticized by many who claim that it is too restrictive (Priest, 1986), yet it has been praised by others for its inclusiveness (Ford, 1986; Staley, 1979a). Priest argues for a broader definition of outdoor education, which includes six major points: 1) a method of learning; 2) an experiential method; 3) occurring primarily in the out-of-doors; 4) multi-sensory; 5) interdisciplinary; and, 6) concerned with relationships (intrapersonal, interpersonal, ecosystemic, and ekistic) (Priest, 1986).

Staley, in a seminal work Outdoor Education for the Whole Child, also states that outdoor education is an informal method of teaching and learning, and Staley supports "education in, for and about the outdoors" as a comprehensive definition of outdoor education (1979a). Ford, suggests that the "in, about and for" definition is the most appropriate for outdoor education in that it tells the place, the topic, and the purpose of outdoor education (1986). Both Staley and Ford take the prepositions in turn and explain how each encompasses the versatility found in outdoor education. Roth, Cantrell, and Bousquet (1980) use the "in, for and about" definition to explain the relationship between residential outdoor education and environmental education, stating: "Environmental education stands not along, not beside, but upon the shoulders of resident outdoor education" (1980, p. 89).

Gilbertson operationally defined outdoor education as "the umbrella under which different types of education occur in the outdoors, about the outdoors and for the natural environment" (1990, p. 17). Gilbertson emphasizes the experiential, interpersonal skill development, and physical aspects of outdoor education. More recently, Knapp (1992) defined outdoor education as "the instructional use of natural and built laboratories beyond the school to expand and enrich learning"(p. 1).
For the current study, outdoor education programs were defined as those which subscribe to the philosophy proposed by Sharp: "That which can best be learned in the out-of-doors through direct experience, dealing with native materials and life situations, should there be learned" (1952, p. 9). This definition was applied as it is comprehensive and well accepted.

The residential aspect of the term residential outdoor education programs must be defined. Technically, any program at which the participants reside overnight could be considered a residential program. However, since many of the social and psychological effects of ROEP have been accredited to the circumstances created by an extended stay and close living arrangements, short, overnight programs have not typically been considered in the same category as those including a prolonged visit. For example, the American Camping Association, reserves the term "resident camp" for those programs having a majority of sessions "which are at least five consecutive days in length" (1993). In their 1984/85 study on the practices of evaluation of ROEP, Chenery and Hammerman operationally defined a ROEP as "an educational program that is conducted at a resident 'outdoor school' (i.e., at a camp, conference ground of environmental center) for a period of at least three days and two nights, and designed to facilitate and enrich learning goals related to the school curriculum" (1985a, p. 35). Based on the definitions provided in the literature, residential programs were defined for this study as those consisting of a visit of at least two nights.

A History of Resident Outdoor Education: Throughout the history of ROEP, several trends in program emphases have evolved. In the 1800's and early 1900's, several organization-sponsored camps (i.e., YMCA - 1885, Boy Scout - 1914 and Girl Scout - 1922) were begun, and the first church sponsored ROEP was begun in 1880 by Reverend George W. Hinkley (Markovits, 1977). These initiatives were influenced by the
nature study movement, social welfare movement, and the first conservation movement. By the summer of 1912, public school camping as a practice had been initiated. The 1930's witnessed an increase in programs sponsored by school districts. The foci of these early camps were societal concerns, "enrichment" of urban youth, and a traditional educational approach emphasizing the three R's (Markovits, 1977). A majority of these summer programs were administered by retired military professionals and teachers (Markovits, 1977).

Concurrent with the public schools' emphasis on education for the whole child (Hammerman & Hammerman, 1968), a new trend in outdoor education began in the thirties and the forties. Catalyzed by the works of Sharp and Kilpatrick, a concentration on the interrelationships of humans and the environment flourished (Markovits, 1977). Sharp also stimulated the research and evaluation work in this field (Markovits, 1977). Hammerman and Hammerman title the period from 1930 to 1939 the "Period of Inception." These authors describe residential outdoor education programs at this time as similar to traditional summer camps in offering nature study, hiking, campcraft, etc., but also as "influenced by the prevailing educational philosophy that the schools should educate for living; consequently, the socialization and work experiences of camp living were stressed" (Hammerman & Hammerman, 1980, p. 4).

The philosophies of outdoor education professionals in the 1950's are well summarized in a statement from Sharp: "the experience of living in the out-of-doors together as a regular part of the school program is not a fad, frill, or extra.... Here the students meet the more subtle problems involved in group living, the problems connected with the unselfish and unbiased consideration of others, the problems involved in fears and prejudices" (1952, p. 21). During this decade, outdoor education
became increasingly accepted by formal educators. According to Hammerman and
Hammerman, by the mid-fifties, "outdoor education was well entrenched in a number
of school systems across the county and, in fact to a certain extent, had become quite
standardized" (1968, p. 82).

The outdoor laboratory was a popular theme in the 1960's, with programs
highlighting the methods of science and the use these methods, i.e., methods of
discovery, inquiry, problem solving (Grilley, 1966). The trend was further enhanced by
the second environmental movement beginning in the 1960's. It was during this
decade, also that the term "Environmental Education" emerged and attention was
devoted to it through the programs.

Writing in the late 1970's, Stailey also suggests that the following trends had
been incorporated in to the outdoor education curriculum:

1. The use of other resources such as the school site and
urban environments to enrich the school's curriculum,
2. Environmental education which emphasizes a
knowledge and concern for environmental quality,
3. Energy education which emphasizes the wise use of
our dwindling energy resources,
4. Adventure education which provides youth with
exciting outdoor encounters and challenges; and
5. Experiential education which emphasizes learning from
one's total life experiences. (1979a, p. 3)

By the 1980's, residential outdoor education was striving to incorporate all
subjects, in an attempt to abandon the idea that only science could be taught in the
out-of-doors. More holistic curricula were presented and marketing toted the
advantages of the inclusive programming available. In the 1990's this trend continues
along with increasing stress placed on environmental education.

In summary, resident outdoor education programs are a collection of distinct
programs with individual curricular priorities. Through certain periods of history,
however, trends in the emphases have emerged (e.g., movement away from wilderness survival skills and toward healthy lifestyle skills) and have lead to the diversity in activities observed in these programs today. Nevertheless, these programs seem to subscribe to the same basic philosophy, as stated by Sharp: "That which can best be learned in the out-of-doors through direct experience, dealing with native materials and life situations, should there be learned"(1952, p. 9).

The Goals and Hypothesized Effects: As a result of the diverse types of ROEP many goals, objectives, and hypothesized effects exist for these programs. According to Wiegman (1980), however, a number of goals seem to be universally accepted. Through personal interviews, this researcher found that camp directors valued "social growth of the students," "personal growth of the students," and "values about the environment" (1980, p. 24). In the same study, through a mailed questionnaire completed by directors, staff, board members, and visiting teachers, Wiegman identified the programs' primary goals as:

A. Social growth, community lifestyle, cooperation, and/or understanding of others is important for the students.
B. Self-awareness, self-esteem, and/or personal growth of students is important.
O. Hands on, first hand, and/or experiential learning are important [sic]
M. The values the students form about the environment are important. (Wiegman, 1980, p. 26-27)

These subjects ranked the following pre-selected goals as the least important:

F. Specific motor skills are important for the students to learn.
K. It is important for the students to learn a specific set of concepts.
J. It is important for the students to learn a body of knowledge.
N. An appreciation of culture and pioneer history is important.
P. Creative leisure time activities are important for students to learn.
C. Resident outdoor education is a ministry of our organization. (Wiegman, 1980, p. 27)
Staley enumerates six goals of general education to which outdoor education contributes. These are:

1. To develop children's knowledge and appreciation of their environment and man's (human's) relationship to this environment.
2. To develop children's understandings, skills and appreciations needed to mentally, emotionally, physically, socially, and spiritually lead productive, creative, and enjoyable lives.
3. To develop children's skills and motivation to learn how to learn (i.e., how to find problems, solve problems, and make decisions).
4. To develop children's ability to evaluate and improve their concepts of personal achievement and self-worth.
5. To develop children's skills and understandings required to work and learn cooperatively with other individuals.
6. To develop the creative potentials within each child.

(1979a, pp. 4-8)

Based on these broad goals and objectives and the specific interests of individual programs, a plethora of benefits has been attributed to residential outdoor education programs. In an article summarizing these benefits, the president for the Institute for Environmental Camping and Outdoor Education and the Foundation for the Advancement of Environmental Education, W.M. Hammerman (1980), placed the results in the following generalized categories: socialization/affective domain, subject matter areas/cognitive domain, and attitudinal development/affective domain. Much research has been conducted on the effects of these programs in the socialization/affective domain. Studies have focused on such topics as self-esteem (Rawson & McIntosh, 1991), locus of control (Freeman, Anderson, Kairrey, & Hunt, 1982), social competence behaviors (Chenery, 1981), perceived conflict (Rider, 1986) and more. In the category of subject matter and cognitive domain, effects have been proposed for multiple subjects and thought processes, including: problem-solving (Bennett & Padalino, 1989), ecological knowledge (Oloke, 1981), and specific health knowledge taught (Santos-Ortiz, 1992). Studies on attitudes towards learning,
teachers, and school in general, toward outdoor recreation (Merritt, 1987), and toward the environment (Raze, 1989), have identified some of the implied positive outcomes in the domain of attitude development/affective domain. However, as previously stated, the actuality of these favorable results remains in question.

**Evaluation of Resident Outdoor Education Programs**

This section will examine literature written about residential outdoor education evaluations through: (1) summarizing accepted rationale behind and purposes for performing these evaluations; (2) exploring the barriers to these evaluations; (3) tracing the development of models, frameworks, or guidelines for residential outdoor education program evaluations; and, (4) presenting evaluation literature from related fields.

**Rationale behind Program Evaluation:** Many authors state that evaluation is a valuable tool to provide marketing information. Henderson (1987) reports that some camp administrators regard evaluation as a method for demonstrating the benefits of the product that they offer. Chenery and Akers (1987) found that through an evaluation of the factors of camp that were important to parents, one camp increased its average return rate from two to three and a half years. Baker writes that evaluation "can be your single most useful sales tool and will yield high dividends" (1976, p. 26) for adventure programs.

Others feel that, although it may be wise to use evaluation information for marketing purposes, this should not be the primary motivation for the evaluation. In the specific field of experiential education, Warner (1984), claims that the generation of propaganda material has been the primary reason for most evaluation plans, resulting in the "relegation of evaluation to a lowly position" (p. 39), in the eyes of camp administrators more interested in programming and staff development. Flor (1991)
adds that marketing should be one of the lesser priorities for conducting evaluation; nevertheless, Flor contends that the information produced by a program evaluation should be used for marketing purposes in addition to its other uses. This author states that this may provide a means for program survival. Through evaluation, programs can demonstrate their credibility and effectiveness in obtaining their objectives and thus can apply this information towards marketing.

Gaining accountability through the establishment of the programs' efficiency is also listed as major reason for evaluation. Parents, teachers, school boards, boards of directors, and sponsoring agencies are interested in the results of the program (Henderson, 1987). In a 1993 article, Henderson and Bialeschki listed accountability as one of the two major reasons for evaluation. (These authors listed decision-making as the other major purpose for evaluation.)

"The bottom line is to improve the quality of learning which students experience as program participants" (Flor, 1991, p. 38). Evaluation for the purpose of program improvement is well supported in the literature. Warner (1984) claims that the key to beneficial evaluations is that they be conducted with the intention to aid in program development, with this will come with the devotion of the camp personnel to produce a valid assessment. In describing "direct evaluation," Nowak (1984) states that evaluation contributes to the program though program justification, evolution, and modification. Evaluation for improvement can include determination of the most effective teaching techniques, evaluation of the staff for individual improvement (Magnuson, 1991), identification of the most valuable activities, assessment of goals, and conclusion of the best "interventions" (i.e., camper groupings) (Henderson, 1987).

Despite the numerous noble purposes for conducting evaluations, the greatest reason evaluations of ROEP are completed is to fulfill a mandate (White, 1975). The
second most frequent reason is in order to produce information to defend funding, and the most infrequent reason is for voluntary program improvement (White, 1975). White states that, when evaluation is mandated, "there is very little hope that the evaluation will produce useful and important information" (1975, p. 198).

The literature (Flor, 1991; Henderson, 1987; Warner, 1984) supports three basic purposes for program evaluation which make it beneficial to the organization. An evaluation can provide a means for 1) program improvement, 2) for proving effectiveness (thus gaining credibility for marketing purposes), and 3) for documenting the ROEP's efficiency (thereby providing information on accountability).

Barriers to ROEP Evaluations: In residential outdoor education, there are barriers which make evaluation of these programs especially demanding. In addition to the typical difficulties of evaluating an educational program, as a result of the often informal nature, affective, behavioral emphases, experiential elements, and uniqueness of each program, evaluators of these programs are faced with intense challenges. Passineau states: "Because many environmental educators feel insecure when it comes to evaluation, little effort has been made to use it to full advantage" (1975, p.384). Henderson and Chenery list specific barriers to ROEP evaluation such as: "Lack of time, hesitancy to intrude on the...experience, and questions about adequate measurements" (1987, p. 30). The following barriers are discussed herein: inadequate time for evaluation, inappropriate models and methodologies, disbelief in the measurability of the outcomes, and untrained evaluators.

Inadequate time: The barrier of inadequate time is well illustrated in the following passage from Chenery and Hammerman:

Resident outdoor education programs proceed a lot like the water cycle: first rain, then run-off, evaporation, and more rain. A group comes in; the program flourishes at an intense pace for 3-5 days; the students and teachers leave; there's a brief respite, then more program to a new
group. In the process we hope that seedlings we have planted do grow. The teachers and parents catch the run-off, and time and future opportunities determine whether that run-off becomes part of new behavior, true learning. While we do want to know more about the "crop" that just passed through our fields, the deluge of the new group forces a focus on the immediate. (1984, p. 35)

Directors of such programs often feel that "all possible effort should be directed toward the campers [students] and their experiences; proof of the value of the experience becomes only secondary" (Henderson, 1987, p. 30). Most administrators understand the value of evaluating (Henderson & Bialeschki, 1984), yet they simply can not allocate adequate time for the development of an in-depth evaluation.

In inappropriate techniques and models: The reliance on inappropriate evaluation methodologies and models is also a significant barrier to evaluations. "A serious threat to the use of appropriate research evaluation methodologies in outdoor education, and in education in general, is that many involved in decision making, funding, and researching such programs have been trained, indoctrinated or led to believe in the superiority of empirical research" (Staley, 1979b, p. 4) In the work, The Research, Evaluation and Measurement Dragons in Outdoor Education, Staley comments that outdoor education programs are not easily fit into the experimental models (i.e., random assignment and rigorous control are not possible), and that poorly conducted quasi-experiments are the results of such effort. This author states that as a result of their faulty design, the quasi-experiments that are conducted on these programs seldom detect significant differences that exist. The persistence of inappropriate evaluation models and methodologies creates a serious philosophical barrier to evaluation of ROEP.

Measurability of effects: Professionals in these fields wonder whether or not the unique effects of these experiences can actually be measured. Educators who
believe in experiential education may not accept traditional quantitative evaluations as a valid measure of the effects of their program. The dominance of this type of evaluation has lead many to believe that evaluation can not produce useful information (Staley, 1979b).

**Inadequate training:** The lack of trained personnel to conduct evaluations is a barrier to effective evaluation. Both Staley (1979b) and Henderson (1987) note that outdoor education professionals' lack of training in the use and understanding of research and evaluation may be one reason for the failures of these evaluations. Effective evaluation requires knowledgeable evaluators. The training of environmental and outdoor educators may be insufficient in providing the necessary knowledge and skills, resulting in unsuccessful or incomplete evaluations.

Thus, the literature indicates that evaluation of residential outdoor education programs have been seriously hindered by the barriers of: lack of time, disbelief in the measurability of effects, inappropriate models and methodologies, and inadequate training. Any attempt to improve the status of evaluation in this field must undertake to lessen these barriers.

**Evaluation and Evaluation Models for ROEP:** Evaluation models of residential outdoor education programs have primarily been derived from evaluation models developed for formal education programs. One of the earliest evaluation models used for education in the United States was that developed by Ralph Tyler for the classic Eight-Year Study. In the words of Worthen and Sanders, "In his [Tyler's] work, evaluation came to be viewed as the process of comparing performance data with clearly specified objectives"(1987,p. 44). The model emphasized the importance of assessing students achievement of behavioral objectives which are based on the programs' goals.
This objectives-based, experimental-type model has become the basis for some ROEP evaluations. For example, Pepper (1952) studied the program, objectives, curriculum, administration, and evaluation of school camping. That study used a traditional goal-based approach to evaluation and collected data from campers, teachers, and parents. Pepper reported that "evaluation is an area of great interest, but one that has not as yet been very extensively or intensively developed" (1952, p. 77).

A variety of evaluation models and techniques have been applied to ROEP. In a 1966 study of the evaluation of 16 residential outdoor education laboratory school programs, Grilley found that:

- 94% of the programs ask participants to fill out questionnaires about their attitudes and interests;
- 94% have participants keep a diary of some nature;
- 94% of the programs are evaluated through the teachers' evaluation of participants' conceptual gains;
- 88% administer evaluation forms pertaining to needed program changes;
- 88% ask parents to respond to a questionnaire;
- 81% conduct pre and post inventories of participants' interests; and,
- 69% administer pre and post sociograms.

The results of this study indicate that formal evaluation of these formal education programs was common.

Gilbertson (1990) writes that formal research on the relationship between residential outdoor experiences (school camps) and environmental attitudes and knowledge did not begin until the 1970's. By 1975, evaluation topics were prevalent in the research literature. From a review of 117 studies of research in outdoor education, D.B. Bennett (1975) found that eight of the studies related to needs assessment; four of these studies dealt with the development of goals and objectives; twenty studies wrestled with defining and improving camp programs and the educational setting; and, fourteen assessed outcomes of the program. Twelve of the studies assessed affective
outcomes; four included some evaluation of cognitive effects; two measured motor
skills; and, none addressed problem-solving or understandings. In reviewing 62
studies from UNESCO, D.B. Bennett found similar emphases: 52 evaluated affective
aspects, 28 addressed cognitive effects, and 13 dealt with psychomotor changes. From
this review, D.B. Bennett concluded that most outdoor education evaluations at the
time used external evaluations.

Many outdoor educators reacted negatively to the use of goal-based evaluation
of ROEP, and authors began suggesting the use of alternative models. In What a Small
Boy and a Hammer Have to Do with Evaluation, White (1975) uses the analogy of a child
playing with a hammer to illustrate the possible negative effects of outdoor education
evaluations which over-use goal-based evaluation models and conventional
techniques (e.g., written tests). As a child would hit objects indiscriminately (with
deleterious effects), so too do evaluators with these tendencies. White suggests several
other models which may be more appropriate: goal-free, adversary, portrayal, and

The first guide of any substantial nature on assessment of residential outdoor
education programs was published in 1975 by the Los Angeles County Superintendent
of Schools. The Guide for Self-Appraisal and Certification of Resident Outdoor
Environmental Education Programs presents an instrument designed to identify the
strengths and weaknesses of individual programs. The instrument consists of a series
of checklists and procedures for preparing for a certification process. This work
provides standards for certification and suggests how a standardized certification
process could be developed. In the appendix, sample evaluation instruments are
included. Another document, the Guide for Self-Appraisal and Certification of ROSS:
Resident Outdoor Science School Programs, was published in 1985.
As evaluation in this field expanded, evaluators realized that the uniqueness of each program necessitated that individualized evaluations be developed for each site. In two 1979 works, Staley urged outdoor educators to realize that "different kinds of programs and goals require different types of research evaluation methodology" (1979b, p. 3). This author described the advantages and disadvantages of two basic approaches (or designs) of evaluation: experimental (which uses quantitative measurement and deductive reasoning in a controlled situation in order to determine causal relationships), and non-experimental (which uses qualitative measurements and inductive reasoning to describe the dynamic interactions of a real situation). Although Staley recognizes that the experimental approach may be useful and can be successful when the system and the variables can be clearly defined and controlled, Staley recommends the non-experimental approach for program evaluations. This design is "ideal for the evaluation of most outdoor education programs because it provided[s] information that is useful for making decisions about program effectiveness. It thus contributes to program-development, not its labeling of success and failures" (Staley, 1979b, p.2).

Staley states that outdoor education evaluators have traditionally been indoctrinated with the experimental approach and have been influenced by decision makers who believe in the superiority of this approach. This author believes that, as a result of this dependence on one evaluation approach, some significant effects of outdoor education may not have been detected in many evaluations.
Urging evaluators to be open to a variety of evaluation approaches, Staley writes that evaluators must convince decision makers of the appropriateness of alternative approaches. Staley provides four statements to guide evaluators in their choice of an evaluation design:

1) Close interaction should be established between evaluators and program designers. The goals, objectives, and the program itself must be clearly described.

2) The evaluator and program designers must agree on the paradigm or perspective from which they will approach the evaluation.

3) A list of appropriate methods should be made with the consideration of the questions of the study and the available resources. And,

4) Selection of the method or methods should be made on the basis of those "most appropriate to answering the questions within the meaning framework (bias) of the decision makers, program designers and evaluators" (1979b, p.4).

In summary, Staley proposes that standardized evaluation criteria and experimental designs are not appropriate for outdoor education programs. Instead, each evaluation should be designed specifically for the needs and goals of the individual program.

The concern for evaluation continued in the 1980's. During this period, Bacon (1982) conducted a survey of school districts which included a resident camping experience in their curriculum. Bacon found that although objective evaluation of the programs was desired by all of the respondents, it was not routinely being conducted (Bacon, 1982). In the early eighties, several authors strove to provide guidance to evaluators through publications. For example, in a 1981 article, "Research/Evaluation of Your Camp Program - It Doesn't Have to be Painful," W.M. Hammerman offered general steps for evaluating a residential outdoor education program. These were: 1) List questions pertaining to the program's impact; 2) Develop tentative hypothesis based on the questions; and, 3) Create an evaluation design (determine
methods, tools, timetables, evaluation staff, and intended use of results). Additionally, during this period, evaluation was incorporated in the education of program administrators. The presence of a unit on evaluation in the American Camping Association-supported *Camp Director Education Curriculum Guide* (Stein, 1981) demonstrates that higher education professionals valued evaluation.

The results of the National Survey of Evaluation Methods and Results in Resident Outdoor education Programs were published in a 1984/85 article by Chenery and Hammerman. These researchers found that practitioners value the results of research and evaluation of their programs; yet, "only a few outdoor centers seemed to be actively engaged in evaluation" (Hammerman & Chenery, 1985b, p. 37). This work also ascertained that the immediate needs of evaluation related to determining the academic, social, psychological, and attitudinal benefits to the students and the identification of the most effective teaching techniques and experiences for attaining these benefits.

Hammerman and Chenery (1985a) report in another article that most ROEP do evaluate; however, they question the effectiveness of these evaluations (Chenery & Hammerman, 1985a). These authors provide specifics on the survey findings. Among the survey respondents, the most commonly applied evaluation methods were: observations of the programs (90.4%), group discussions (78.5%), individual discussions (74.8%), and written survey questionnaires (73.3%). For the majority of the respondents' evaluations, classroom teachers and participants completed some form of evaluation. Typically, administrators determine the content of and conduct the evaluation. The results indicated that program or staff improvement was the prevailing purpose for the evaluations (1985, pp. 36-38). Additionally, respondents submitted numerous example instruments of evaluation of the program or staff.
Chenery and Hammerman reemphasize Staley's concerns for evaluation. "An evaluation system for an outdoor education program or center must be based on that program's particular goals and objectives as well as the center's specific needs for information" (Chenery & Hammerman, 1985a, p. 40). These authors support the use of multiple methods, each chosen specifically for its ability to obtain useful information. They also underscore the importance of detailed descriptions and advise evaluators to reject experimental designs.

Chenery and Hammerman present an objectives-based evaluation framework entitled "The Evaluation Development Process" as well as examples for each category within the framework. The framework includes: Program Objectives; Signs of Accomplishment of Objectives; Activities/Program Related to Objectives; Evaluation Questions; Decisions to be Made on the Basis of this Evaluation; Actors Involved; Evaluation Methods: What? Who? When?; Evaluation Product or Summary; Distribution or Use of Evaluation Summary; and, Costs of the Evaluation (1985a, pp. 40-41). This framework is one of the most comprehensive, contemporary, objectives-based guides to evaluation which is specifically designed for residential outdoor education programs. Chenery and Hammerman suggest the need for an established model to guide evaluations of ROEP.

The national concern in the 1980's for educational accountability also affected outdoor education. An article by Baker describes an accountability model as a form of evaluation of outdoor education programs. Baker identifies six steps to accountability: "(1) needs assessment; (2) prioritize goals from assessment data; (3) develop program goals; (4) develop measurable program objectives; (5) develop specific instructional activity objectives; and, (6) plan an evaluation" (1983, p.13). This model is goal-based and is designed specifically to produce information for decision makers.
Both responsive evaluation and naturalistic inquiry have been used in general educational evaluation. These approaches have been successfully used to evaluate residential outdoor education programs. Chenery and Russell (1987) describe such an evaluation and offer reasons why this approach is appropriate for ROEP. These authors dispute the claim that "the choice of an evaluation approach should be based on conditions in a setting that make it favorable to one approach or another (such as size, nature of the program, and availability of resources)" (1987, p. 31). Instead, Chenery and Russell state, "the choice should arise from a decision about the philosophical foundation" (1987, p.31). Their article supports responsive evaluation as it "provides information that is grounded in experience and offers a method to match the richness and wonder of leisure experience" (1987, p. 32). They also describe the nature of responsive evaluation and naturalistic inquiries, providing brief descriptions of: qualitative methods, purposive sampling, use of the natural setting, evaluator as the data collection instrument, grounded theory, emergent design, etc. This article exemplifies how evaluation theories from general education can be tailored to the field of residential outdoor education.

The effort to provide program directors with guidance was continued most recently by Henderson and Bialeschki. In their 1993 article "Camp was Great, but the Water was too Cold," they assert the importance of evaluation as an integral part of the camp program and state that "the major reasons evaluations are not systematically conducted in camps are because camp personnel do not know how to set up an effective evaluation process, how to analyze the data, and/or how to interpret the data in useful ways to assist in accountability and decision-making" (1993, p.32).
The framework that Henderson and Bialeschki (1993, p. 32) present, "A Matrix for Camp Evaluation," involves four fundamental components: staff, program, facilities, and administration which may each be evaluated on the basis of objectives, standards, and outcomes. The authors also provide basic definitions and descriptions of evaluation terminology (i.e., formative evaluation, qualitative data, random sample, etc.) and of a small number of evaluation tools (i.e., Likert scales). This article provides a brief summary of evaluation in non-technical terms for camp administrators.

**Evaluation Models from Related Fields:** Evaluation models found in the literature of related fields may be adaptable for use by residential outdoor education programs. More formalized evaluation models have emerged within fields such as extension and environmental education. This section will present some of the leading models in each of these fields.

**Evaluation of Extension programs:** An evaluation model by C. Bennett (1975) has dominated the evaluation of extension education programs. In "Up the Hierarchy," C. Bennett outlines an objectives-based model which uses a seven-tiered hierarchy as a framework for understanding different levels of evaluation. [See Figure 1.]
* "KASA change" means changes in knowledge, attitudes, skills, or aspirations to change behaviors.

FIGURE 1. Bennett’s Hierarchy of Evidence For Program Evaluation.
The first level, inputs, includes the projected allotments of time, money, and other resources to the program. For the second level, activities, the actual plans for the program, the curriculum, etc. are the focus. People involvement, the third stage, addresses the number and demographic characteristics of persons who are affected by the program. The participants' reactions (emotional, interest level, etc.) to the program activities and presentation are assessed at the fourth level (note, this does not include reactions to the content of the program). The content of the program is evaluated through the fifth level, KASA change. KASA change represents changes in knowledge, attitudes, skills, and aspirations to act in a certain manner. Actual modifications of behavior are assessed at the sixth level, that of practice change. Evaluations at the highest level, end results, are concerned with the programs effects on a societal level.

C. Bennett explains that "Evidence of program impact becomes stronger as the hierarchy is ascended.... The difficulty and cost of obtaining evidence of accomplishments increases as the hierarchy is ascended.... While hard evidence is usually ideal, it's more expensive and difficult to obtain [italics removed]" (1975, pp. 9-11). Also presented is a brief guide to selecting evidence for the evaluation of a program. Books such as The Answers to Program Evaluation: A Workbook (Spiegel & Leeds, 1992) rely solely on this model. This demonstrates the overwhelming acceptance of this model by the Extension Service programs. C. Bennett's hierarchy provides a framework for the design of objectives-based evaluation of educational programs and may be applicable to evaluations of residential outdoor education programs.

Additionally the Extension Service has provided education about evaluation through in-service, training workshops. Van Tilburg and Archer present a model program which prepares an Evaluation Resource Task Group through focusing on the
topics: "data collection and analysis, formative and summative evaluation, interpretation and communication of results, evaluation management, and consultation" (1986, p. 1). Educational programs about the evaluation of extension programs may provide a prototype for models of evaluation for residential outdoor education programs.

*Evaluation of environmental education programs:* Environmental education is often a recognized curricular element of formal education programs. As such, many environmental education programs have been subject to mandatory, summative evaluations; thus, formal evaluation models have arisen within the environmental education literature. By definition, outdoor education includes some form of environmental education (education for and about the environment). As Wiegman found in a 1980 study, a majority of the camp directors who were interviewed reported that one of their program's primary goals was to improve values concerning the environment. The literature supports the connection between outdoor education and environmental education programs, and examples can be found in which an evaluation of an "environmental education program" was conducted at a residential outdoor education program (e.g., Dayton, T.G. & Allen, R. 1984). Thus, some environmental education evaluation models may be appropriate for evaluations of ROEP.

"Evaluating Environmental Education Programs," a chapter in *Environmental Education Strategies Toward a More Livable Future*, was written by D.B. Bennett (1974), one of the leading authors in evaluation of environmental education programs. Bennett draws on the field of psychology to develop an evaluation model which follows an experimental approach. The author states that the goals of environmental education consist of cognitive, affective, and skill areas of individuals' attitudes,
behaviors, and values, and that affective and cognitive components merge to form attitudes which cluster together to create values (Bennett, D.B., 1974). The basic tenets for D.B. Bennett's evaluation model are: (1) the educational experience is the independent variable, and (2) the participants' ecological value system and their ability to act based on that value system are the dependent variables. D.B. Bennett emphasizes the need for basic behavioral objectives for which achievement can be measured. From these basic objectives, D.B. Bennett proposes that three types of specific objectives be developed. One specific objective should be developed for each of the following evaluation techniques:

1. process evaluation-observation and evaluation during the learning activities,
2. formal evaluation - primarily pencil-and-paper instruments for pretesting and posttesting, and
3. unobtrusive evaluation-techniques of data-gathering during which students are unaware of being observed and evaluated. (Bennett, D.B., 1974, p.134).

D.B. Bennett also presents an example of how the evaluation model was used to evaluate a junior high school's environmental education curriculum. This example illustrates the behavioral objectives and the three types of techniques. This 1974 work provides an overview of environmental education evaluation, an experimental, behavioral objectives-based model, information on evaluation design issues, and a detailed example of the application of the model to a program evaluation.

In 1984, D.B. Bennett produced a thorough exposition entitled Evaluating Environmental Education in Schools: A Practical Guide for Teachers. Environmental Education Series 12. This guide provides practitioners step-by-step instructions for the entire evaluation process. D.B. Bennett describes evaluation as central to the educational process which includes needs, goals and objectives, instructional activities, and student learning. This author promotes evaluation as it leads to
program improvement, growth in student learning, and a better environment (Bennett, 1984). This guide recommends that formal evaluations be based on quantifiable behavioral objectives, instructs the reader in experimental designs, and primarily presents quantitative techniques; however, D.B. Bennett does recognize the importance of qualitative data and presents succinct sections on qualitative techniques such as unobtrusive observation.

D.B. Bennett addresses the following major topics: (1) definitions of and rationale for evaluation; (2) choice of evaluation object; (3) evaluation question, design, and instruments; (4) data collection (descriptions of techniques), analysis and interpretation; (5) use of results; and, (6) an illustrative case study. Each section provides detailed descriptions and practical advice for novice evaluators. For example, in describing each technique, the author includes an overview of the benefits and limitations, useful guidelines for development, and specific examples. In summary, D.B. Bennett was successful in creating a practical guide for formal environmental education evaluation; however, this guide essentially limits evaluation to an experimental approach.

In a chapter of McInnis and Albrecht's book What Makes Education Environmental?, Passineau advanced a similar evaluation philosophy, emphasizing the need for a reductionistic approach to describe the "unique and non-unique features of EE [environmental education]" (Passineau, 1975, p. 379) in order to discover causal relationships. Passineau encourages the development of scientific rigor and appropriate measurement tools for environmental education evaluations. After describing an informal evaluation approach ("The Question Model") which uses simple questions (Why are we evaluating? What are we evaluating? Who are we evaluating? etc.) to guide the evaluation, Passineau proposes that there are two basic types of
formal educational evaluation models: (1) those which focus on curriculum development (usually a goal-based approach), and (2) those which assess the entire program (usually a systems approach). Passineau dismisses the curriculum development models as inappropriate for environmental education evaluation as they are not comprehensive; Passineau details several systems approaches; and, this author presents "The Evaluation Model" which was created by blending several existing models.

"The Evaluation Model" describes three phases of evaluation: (1) program planning, (2) implementation, and (3) final evaluation. This model suggests that the evaluation and development of a program should occur simultaneously and that evaluations should first include a description of the development. Passineau proposes general guidelines for evaluation, stressing the importance of determining and measuring the achievement of specific goals and educational objectives.

In addressing evaluation methods, Passineau claims that evaluators should take advantage of the numerous techniques which have been developed in the field of educational evaluation, and that multiple techniques are probably a necessity for a comprehensive formative and summative evaluation. The author writes, "Although many techniques exist which could be used to assess environmental education efforts, many are not being used because evaluators seek techniques specifically designed for EE [environmental education] when more general ones would work equally well, such as questionnaires, sociometry and behavioral observations" (Passineau, 1975, p. 340). In addition to these, Passineau recommends unobtrusive measures for a valid assessment of attitudes, values, and behaviors which are especially susceptible to social influence. Because these programs are diverse in approach and topics, this author dismisses standardized tests as limited in value for environmental education.
In agreement with Bennett, Passineau (1975) supports the use of behavioral objectives. Passineau proposes:

The view that behavioral objectives do not promote the entire realm of educational potential including decision making and creativity is likely based upon the realization that behavioral objectives are much easier to formulate for the lower levels of cognitive learning than for the higher levels. In theory, however, behavioral objectives should apply equally well for the higher levels of cognitive learning and the affective domain. (Passineau, 1975, p. 394)

Passineau does, however, encourage the assessment of unintended effects in addition to that of intended effects. In summary, Passineau presents an evaluation model adapted for environmental education and a theoretical, academic account of several evaluation models as they apply to environmental education.

Robottom (1989) proposes that the fundamental principle of environmental education is that it is "socially critical education." Environmental education challenges students to think critically about value-laden issues and encourages a close association between knowledge and action. Robottom criticizes "technicist" (those following a reductionistic paradigm) evaluations for their inability to deal with value questions, stating, "technicist thinking may actually be contrary to environmental education's central notion..." (p. 442). Robottom chooses a "critical" approach to evaluation which would employ such methods as: illuminative evaluation, phenomenology, and case studies (1989).

One of the cardinal characteristics of environmental education is its emphasis on all three domains of knowledge: cognitive, behavioral, and affective. "Sensing and feeling" activities have been an essential aspect of outdoor and environmental education. Experiential and discovery learning, dominating environmental education, exercise the affective domain. Since these programs are concerned with the manipulation of emotions, values, ethics, and beliefs, all within the affective domain,
some authors feel that a thorough evaluation of these programs must address the program's effect on the participants' affective domain. Emotions, values, ethics, and beliefs are complex, intensely personal constructs sensitive to the subjective reality of the human experience. Clacherty and Ballantyne suggest that a phenomenological approach to environmental education evaluation is necessary in order to "provide a window into 'the inside'" (1990, p.32). This inside view could be the key to comprehending these complex personal realities.

Recently, another model for evaluation of environmental education programs has been developed in Italy. Mayer proposes evaluation through an indicator system. Ideally, a system of indicators measures distinct components of the system of interest, but also provides information about how the individual components work together to produce the overall effect. In other words the whole of the information to be gained from a system of indicators is greater than the sum of its parts." (Oakes quoted in Mayer, 1991, p. 333).

Mayer developed three sets of quality indicators upon which environmental education programs can be evaluated. The first indicator is "the relevance and concreteness of the project in the local situation" (1991, p. 334). The second indicator, school innovation, pertains to the relationships among those involved with the program as their value systems change as a result of the program (i.e., improved teaching methods, joint research among participants and leaders, etc.). The third indicator "analyses the path of change developed by the project and by processes that the project implements" (Mayer, 1991, p. 334). These indicator groups were used to create a qualitative questionnaire to be completed by teachers in order to assess the quality of environmental education projects. Mayer reports that: the indicator system is valid and reliable; it has enabled Italy to create a cognitive map of its environmental education projects; and, it aids in the identification of potential areas for future improvement.
Through an analysis of the literature on the evaluation of residential outdoor education as well as that of related fields, several evaluation models and sets of guidelines have been identified. A majority of these models are goal-based and have been adapted from general, environmental, or extension education programs. Repeatedly, authors have called for more research in this area and for the development of more extensive models or guides for use by practitioners. Additionally, many have articulated the importance of non-standardized evaluations for each unique program.

**Evaluation Literature Summary**

**Overview:** Evaluation as a formal practice is not new. In fact, there are records of the Emperor of China in 2200 BC ordering formal proficiency testing of public officials. Through the many years that humans have been interested in assessment, a myriad of purposes for, forms of, and uses of evaluation have been created. This term, "evaluation," is commonly used in schools to describe the process of determining the information which individual students have gained; the same term may apply to the process of examining a piece of art. For the purposes of this study, the definition of the Joint Committee on the Standards for Educational Evaluation will be adopted for its encompassing nature. Evaluation is: "a systematic investigation of the worth or merit of some object" (1981, p.12).

The use of a carefully planned design is a critical part of any evaluation. Sanders and Nafziger (1975) write that evaluation design is important in insuring open communication about the project, providing direction, assuring completeness of the evaluation, and promoting efficiency through organization. To guide the planning of evaluations, numerous "models" of evaluation design have been created. In "The Status of Research in Models of Product Development and Evaluation," Smith and
Murray summarize the research base supporting evaluation models of instructional products. They define such models as "simply designs of exemplary processes or procedures" (1974, p. 2) as opposed to theoretical models which attempt to explain or predict. These authors support the use of the term "model" for this type of heuristic device which may be "used to guide evaluation design where there are several competing conceptualizations of appropriate evaluation procedures, each based on differing assumptions and values concerning the evaluation process itself" (1974, p. 4).

In addition to their concern over (1) the lack of a clear definition of, (2) the purposes for, and (3) the use of these models, Sanders and Nafziger indicate the need for further research on the utility of various models.

**Evaluating the Model:** Sparse research on the assessment of evaluation models can be found in the literature on meta-evaluation and, to some extent, in the work of researchers appraising their own models. Lacking guidance from the literature, the developers of such models typically have created their own assessment approaches, often using case studies of one program's application of the model (Bennett, 1974; Jacobson, 1990).

In appraising a system for the design of instructional materials, Parker, (1973) considered the following variables: the clarity of the language, degree of training in use of the model, problems with the procedures, concern for the effects of the model on outcomes, user-effectiveness, and user-efficiency. To analyze the adequacy of evaluation designs, Sanders and Nafziger propose that the following topics be addressed: scope, relevance, flexibility, feasibility, replicability, objectivity, representativeness, timeliness, persuasiveness, protocol, and ethical considerations (1975). In 1981, the Joint Committee on Standards for Educational Evaluation published a guide to the standards for educational evaluation. This national group of
scholars recommends that assessments of evaluations should examine the topics of utility, feasibility, propriety, and accuracy. Ratner (1991) assessed environmental education materials by considering the usefulness and appropriateness of the content, design, and writing style.

All of these works attempt to determine the degree to which their models are useful by addressing separate concerns within the general construct of the model's utility. The Joint Committee on Standards for Educational Evaluation, defined the utility of an evaluation as: "The extent to which an evaluation produces and disseminates reports that inform relevant audiences and have beneficial impact on their work" (1981, p. 156). In 1988, Mullins and Watson, in assessing environmental education (specifically interpretation) materials, define utility as "the item's usefulness to you in carrying out your job duties" (p. 35).

From this literature review, it can be established that 1) there is a need for the evaluation of ROEP; 2) identifiable barriers to evaluation of these programs exist; 3) successful evaluation models and guides must address these barriers; 4) attempts have been made to provide evaluation models or guides for residential outdoor education and similar programs; and, 5) the over-arching concept concerning the assessment of evaluation models is that of the model's utility.
CHAPTER III

PROCEDURES

Research Design

The study design consisted of four stages: (1) a review of the literature that indicate the need for a guidebook; (2) the drafting of the guidebook; (3) an initial assessment of the guidebook; (4) a comprehensive assessment of the guidebook; and, (5) the preparation of the prepublication draft of the guidebook (see Appendix K). An overview of each of these steps follows.

Review of the literature: The initial step for this research process was to review relevant literature in order to identify evaluation techniques which are applicable to ROEP. Literature available through The Ohio State University and the Lorado Taft Campus, University of Northern Illinois, libraries and electronic search systems addressing camping, environmental education, environmental management education, experiential education, or outdoor education program evaluations was studied. Through this literature review, information was gathered on numerous evaluation models, barriers to evaluation, purposes of ROEP evaluations, decision steps involved in evaluation design, and evaluation techniques. Eight evaluation models were selected for inclusion in the guidebook. For each model, both information on general design, benefits and weaknesses, and examples of the application of the model to ROEP were obtained. Five decision steps were reviewed. Seventeen techniques were found to be
appropriate for ROEP. These techniques were categorized according to purpose, evaluation design, or model, sources of data, and resources required.

Development of the draft: In the second stage of the research, a heuristic model was developed to describe the purposes, requirements, and general types of outcome for each of the evaluation techniques. This information was organized into a decision framework. The framework, or decision tree, was designed to facilitate the selection of a technique by ROEP administrators, based on their evaluation audiences' needs, their purposes for evaluation, the available sources of data, an appropriate time frame, and available resources.

The guidebook was designed to explain the decision-tree framework. The initial draft included text material with limited emphasis on the page layout and little illustrative material. At this stage, the guidebook included: 1) introductory information on the value of evaluation, definitions of program evaluation, and an explanation of the steps in the design process; 2) individual chapters addressing each of the decision steps, followed by worksheets; and, 3) brief descriptions of appropriate evaluation techniques. The decision steps which were included were: audience, purpose, design, sources of data, time frame, resources, and reliability and validity. The worksheets were designed to help the reader to apply the principles that were presented in the chapter to the evaluation of a specific program.

Following the decision step chapters, sixteen techniques were presented. Several of these were grouped together into general sections (e.g., focus groups and personal interviews were explained as verbal techniques). For most techniques the following information was provided: a definition, a brief explanation of the procedure, some benefits and weaknesses, and references for further information. Three appendices were included to provide additional information on sampling, non-
response error, and hints for successful evaluations. During the development phase, an emphasis was placed on the ease of use, applicability, feasibility, and projected utility of the draft. Additionally, the instruments (in the form of a written questionnaire) for external assessment of the guidebook were developed.

**Initial assessment:** An initial external assessment of the draft of *A Guidebook for Evaluating Residential Outdoor Education Programs (R.O.P.E.)*, was conducted during the third stage of the project. Five expert program directors and five graduate students who have experience in the area of ROEP were asked to review the draft. (See Appendix I for a listing of the reviewers.) These reviewers served as the pilot and field test participants, respectively, in order to determine the reliability and validity of the instruments. (Further details concerning the pilot and field test are included in the Instrumentation section.) Both written and verbal responses from these reviewers were used to revise the draft and to improve the survey design.

Most of the comments from the pilot and field test participants focused on the presentation of the material (i.e., corrections of mechanical errors, suggestions for improving the organization of the material, or remarks about the page layout, spacing, etc.). Multiple comments concerning several of the tables confirmed that the tables were confusing to the readers. Four of the graduate student reviewers indicated that the text was too academic and expressed concern that practitioners would not make use to the guidebook.

The only extensive comment which addressed the content of the model came from one of the program administrators who suggested an additional model that could be included. A majority of the pilot and field test respondents wrote that they felt the model was a comprehensive presentation of the content material.
The changes which resulted from the initial assessment were primarily mechanical, organizational, or page formatting. Mechanical changes included: correcting typographical errors, spelling, and grammatical mistakes; developing conventions for capitalization and hyphenations; and, removing gender specific pronouns. The movement of the validity and reliability section to an appendix and the creation of another appendix for description of data collection times for quasi-experimental evaluations were the major organizational changes. Examples of the formatting changes included: placing an icon on the first page of each chapter to illustrate the placement of the chapter within the entire text, adopting uniform organizational headings, and adding a header on every page. Additionally, the lay out and captions for two tables were improved. Major additions were: a table of figures, one table to the "sources of data" section, a worksheet to the "resources" section, and one model to the "design" section. Finally, the figure on the cover of the guidebook was revised to illustrate the sections of the guidebook as opposed to the stages of a traditional evaluation. The draft was then prepared for the next stage of research, the comprehensive assessment.

Comprehensive assessment: In this stage of the research, the guidebook was reviewed by two panels of experts. One panel was selected from the authors of literature relating to evaluation of residential outdoor education programs, and thus consisted of academic-oriented experts (n=11). These panelists were asked to review the guidebook, to comment on it's content, and to complete a questionnaire. (For a detailed explanation of the questionnaires, see the instrumentation section and Appendices E and G). The other panel was comprised of persons who are administrators of ROEP (n=9). This panel of field experts was asked to review the
guidebook, to provide comments on the guidebook, and to respond to a mailed questionnaire pertaining to the presentation of and the content provided in the guidebook.

The following is an outline of the procedures which were utilized for the data collection. First, the subjects received a letter which acquainted them with the research project and requested their participation (Appendix C). This mailing included a postcard to be returned indicating acceptance or rejection of the offer to participate. Participants were then mailed a full packet containing:

1. a cover letter with reviewing instructions (Appendix D or F);
2. the draft guidebook (Appendix K);
3. a stamped, self-addressed envelope;
4. a questionnaire (Appendices E or G);
5. a red pen; and,
6. a postcard to be returned independently in order to indicate permission to include the reviewer's name in the published guidebook.

The panelists were requested to return the questionnaire and the pages of the guidebook upon which they had commented by a specified date two weeks after they had received the packet. A code number on the back of the questionnaire was included for the purpose of identifying non-respondents. Approximately one week after the packets were sent, a follow-up postcard was sent to each of the panelists to ensure their receipt of the packet (Appendix H).

Since the intent of the study is not to generalize to the entire population of camp directors, lack of responses did not generate the traditional "non-response" error; therefore, no additional resources were spent to gain responses from the non-respondents. Data from the accepting sample was deemed appropriate for use in the study.
Preparation of the prepublication draft: In the final stage of this research process, feedback from the panelists was analyzed and used to revise the content of the guidebook. Both the qualitative and the quantitative data were incorporated into the revision process. While the quantitative data provided general information, many of the qualitative suggestions were directly included into the guidebook.

Subject Selection

For this study, two independent populations were surveyed. As indicated by previous research (Grilley, 1966), appropriate populations to assess an evaluative instrument for outdoor education programs include directors of the camps and selected highly educated professionals in outdoor education. The fourth stage of the research was conducted with two panels of experts in the fields of outdoor education, evaluation, and residential programs, or any combination of these.

A list of perspective academic experts for a panel of academic experts (the accessible population, n=21) was developed through a review of the literature. The target population for the panel of field experts was administrators of residential outdoor education programs in the United States. [Note: the terms "field expert" and "program administrator" are used interchangeably in this text.] The accessible population was approximately 1,500 camp directors whose residential camps are accredited by the ACA in 1993. Thirty perspective field expert panelists were purposively selected to represent a wide variety of programs. Variables considered in the selection process were: location, program emphasis (therapeutic, private, environmental education, organizational-based, religious, adventure, etc.), clientele (age, gender, socioeconomic status), and size of program. Participation in the study was solicited through written correspondences (see Appendix C) and was strictly voluntary.
It is acknowledged that the population from which the sample was directly drawn differed from the total population of ROEP administrators. One of the critical differences between these populations results from programs that did not apply to or were not accepted by the American Camping Association (ACA) for accreditation. The ACA is a large nationally known organization to which many residential programs apply for accreditation; therefore, this list of programs represents the most comprehensive listing of the residential outdoor education programs in the United States. Inaccuracies in the ACA listing to be used, which would result in frame error, were minimized by obtaining the most recent records.

**Outcome Measures**

For this study, the major construct of concern was the models' perceived utility. Two different populations were studied in order to obtain a comprehensive assessment of the guidebook in terms of theoretical soundness and practical application. The issues about evaluating evaluation models (i.e., meta-evaluation) or guidebooks addressed by the previously mentioned authors (JCSEE, 1981; Reese, 1993; Sanders & Nafziger, 1975; Ratner, 1991) can be categorized as concerns about either the presentation, the content, or the overall performance of the guidebook; therefore, subject's opinion/reasoned professional judgment about the presentation, the content, and the performance of the model were the major variables studied. It should be noted that these variables are opinions.

The definition of an "opinion" has been considerably debated in the context of the distinction between this term and attitude. Thurstone (1970) defined opinion as simply the verbal statement of an attitude, and stated that attitude denotes "the sum total of a man's inclinations and feelings, prejudice or bias, preconceived notions,
ideas, fears, threats, and convictions about a specified topic" (p. 128). However, Rosenberg and Hovland (1960) present a three-component view of attitude, claiming that the concept of attitude is composed of affect, cognition, and behavior.

Fishbein and Ajzen (1975) suggest that the critical aspect of an "attitude" is its evaluative or affective nature. They state that the term "opinion" is applied to the measurement of the evaluative aspect of attitude, that opinions have not been proven to follow different scientific laws than attitudes and, that therefore, a distinction should not be made between the terms "opinion" and "attitude." More recently, The Encyclopedic Dictionary of Psychology (1991) defines opinion as "a belief held by a person about other people, objects, or events. Opinions are generally narrower than attitudes and are usually based on a factual item of information" (p. 201).

For the current study, Thurstone's definition of opinion as the verbal expression of an attitude was used. The term "reasoned professional judgment" has also been included to reflect that the attitude of each respondent was solicited under the assumption as a professional person, the respondent would provide a reasoned response based on their expertise or experiences in the field. This attitude was operationalized through the use of written responses to scales. This study focused on the evaluative aspects of the subjects' attitudes toward the model's utility, as emphasized in the works of Fishbein and Ajzen.

Variable 1: Variable One was defined as the subject's opinion or reasoned professional judgment toward the presentation style. This included an evaluation of the semantics and the syntax. "Semantics refers to the symbols or elements themselves" (Ratner, 1991, p. 45). The semantics category assessed the appropriateness of the vocabulary and writing style. "Syntax refers to how the elements are arranged and organized" (Ratner, 1991, p. 46). This category identified
problems with the flow of the information, the cover, the clipboard pages, and the page layout. Variable One was operationalized through the use of a six item, written semantic differential instrument. The panelist's open-ended comments provided on the questionnaire and in the guidebook itself were analyzed to provide additional information to operationalize this variable.

**Variable 2:** The second variable addressed in this study was the panelist's opinion or reasoned professional judgment of the content of the guidebook. The constitutive definition of the user's opinion of the model's content was concerns that were expressed about the model's content and face validity. "Content validity generally refers to the representativeness of the instrument content to the topic being measured" (McCaslin, et al., 1979, p.17). Another researcher, Kerlinger, refers to content validity as "the representativeness or sampling adequacy of the content - the substance, the matter, the topic of a measuring instrument" (1986, p. 417). This variable encompasses the relevance of the material to the subject's interests, the objectivity, and the scope of the information provided.

The operational definition of this variable was the panelist's scores on a two part written numerical rating scale (see Appendix E or G) which addressed the quantity and the quality of twelve topics and open-ended written comments which were provided on the questionnaire and throughout the guidebook.

**Variable 3:** The third variable was defined as the panelist's reasoned professional judgment or opinion of the potential performance of the guidebook. The constitutive definition of this variable included the panelist's concerns of objectivity, feasibility, usefulness, flexibility, and efficiency. The operational definition of this variable was the panelist's mean score on a five item numerical rating scale and their open-ended comments on the questionnaire and on the guidebook itself.
**Rationale:** By addressing these three categories, content, presentation, and performance, the issues which were raised in the literature about evaluating models, designs, and evaluations, will be included in this study. The content variable treated Sanders and Nafziger's questions of relevance, scope, flexibility, objectivity, ethical considerations, representativeness, and accuracy; JCSEE's standards of feasibility, propriety, and accuracy; and, Ratner's inquires on the usefulness and appropriateness of the content. For determining opinions about the presentation, Ratner (1991) recommended the establishment of the subcategories, syntax and semantics. Together, these allow the presentation category to cover flexibility, timeliness, persuasiveness, feasibility, design, and writing style. By recording opinions concerning the overall performance, issues such as relevance, objectivity, representativeness, and feasibility were addressed. The most appropriate source of information on how useful the model is to ROEP administrators could be obtained through direct feedback from the administrators and others currently in this field of research. Therefore, the analysis of the subjects' perceptions of the model's presentation, content, and performance was a valid measure for exploring the utility of this model.

**Instrumentation**

**Description:** The primary data collection procedure used for this study was that of the mailed questionnaire. Issues which were taken into consideration for the choice of data collection method for this study were: limited personnel, minimal budget provided, likelihood of researcher influence on the participants' response, previously recorded high response rates from the population, and anticipated time requirement of the subjects for analyses of the guidebook. One of the primary advantages of use of the mailed questionnaire for this study was that it allowed the respondents to review
the guidebook at their own pace and to critique it in an unhurried manner. It was important that the panelists were allowed the time to consider each question extensively in order to minimize superficial responses. This technique also allowed the respondents to consult with others, an attribute listed by Dillman (1978) as a disadvantage of the mail survey. For the research reported herein, however, this possibility was an advantage because the primary concern was that of obtaining information for the improvement of the model, not that of comparing the individual subject's responses.

The mailed questionnaire methodology was also chosen for the likelihood that researcher influence would be minimized. To assist in the revision process, it was critical that the respondents provided their honest feelings about the guidebook. Other possible methods of data collection, such as face-to-face interviews, might have jeopardized the validity of the responses since the subjects would have known that the interviewer was one of the authors of the guidebook. In such a case, the subjects might not have volunteered their criticisms of the work in the presence of the author, and thus they might have provided socially acceptable, falsely positive responses.

The mailed survey technique was appropriate for this research project, considering the restrictions of low budget and minimal available personnel. The concern over non-response error, a noted disadvantage of mailed questionnaires, was assuaged through an examination of the minimal non-response problems found in previous work with this population. Traditionally, the subjects are well-educated, innately interested in improvement of their field through research and extremely responsive to research projects. In a study, including multiple mailed questionnaires to a population identified as experts in the camping field, Henderson and others (1983) obtained an eighty percent response rate on the initial survey, an eighty-five
percent response on the second survey, and a ninety-five percent response on the final round of questionnaires. Therefore, it was felt that individual item non-response and overall refusal rates would be low and did not discount the mail survey as an adequate method of data collection for this population.

However, the possibility for non-response error was not be negated. A comprehensive plan for pre-contact and follow up was implemented. In addition, preventive measures against non-response were taken through the use of a modification of Dillman's Total Design Method (1978).

Above, non-response error and errors in the choice of method have been discussed. These sections have addressed the primary error concerns; however, there was the possibility of intentional or unintentional misuse of the instruments which could have caused error. The potential for unintentional misuse was reduced through thorough documentation of the process of instrument development and data collection. The probability of intentional misuse of the instruments by the subjects was expected to be small. As stated above, based on the return rates of previous studies, it was expected that this population would be very receptive to the study, and therefore would not misuse the instruments.

**Instrument Development**

Two questionnaires were developed for the two panels. (See Appendix E for the questionnaire that was sent to the academic panelists and Appendix G for the questionnaire sent to the field expert panelists.) Although these questionnaires contained a majority of the same items, the wording of the directions and a few of the items differed. Each questionnaire contained: 1) a six item instrument for
determination of the opinion of the presentation; 2) a twelve item instrument for operationalizing the respondent's opinion of the content; and, 3) a five item instrument used to quantify the respondent's opinion of the guidebook's overall performance. The questionnaire that was completed by the academic experts also include two open-ended questions for that group only: "What other information do you feel should be included?" and "What should be removed?" The questionnaire which was completed by the program administrators included two close-ended questions for that group only about their interest in buying the guidebook and one question asking for descriptive information about their program.

Instrument 1: The first instrument was developed to measure the subject's opinion of the presentation of the material in the guidebook. This variable specifically addressed the panelist's affect towards the item; thus it was appropriate to use a semantic differential scale. The following six categories within the realm of the presentation were identified to be included as items: vocabulary, page layout, cover, writing style, clipboard pages, and information flow. A list of eight bipolar adjective pairs for each item was compiled from pairs which ranked positively on the evaluative factor of a factor analysis conducted by Osgood, Suci, and Tannenbaum (1957) and from pairs generated by the author. (See the instrument in Appendix E or G.) Space for additional comments was provided.

Issues of error common to this type of instrument include face validity and response sets. When the word pairs were originally chosen, the degree to which they appeared to be appropriate measures of the variable were considered in order to enhance the face validity. The issue of face validity is further discussed in the validity assessment section. To minimize the effect of response sets, the positions of the positive and negative adjectives were varied.
Instrument 2: The second instrument was developed to measure the opinion of the model's content. In accordance with the specified definition of "opinion" as the expression of an attitude, a measurement technique for attitude was determined to be appropriate to assess this variable. Instrument types designed for attitudinal variables were reviewed, and it was determined that a numerical rating scale could be utilized to record the respondent's opinion.

Instrument 2 was a twelve item, double-scaled rating scale (included in the questionnaires in Appendices E and G). Each item was either a distinct section of the guidebook or a feature which appeared throughout the guidebook. The items included were: 1) evaluation audience, 2) purpose, 3) sources of data, 4) evaluation design, 5) resources, 6) time frame, 7) techniques, 8) glossary, 9) appendices, 10) further references, 11) pictures/graphics, and 12) examples. These scales assessed the respondent's opinion of the quantity and of the quality of each of these topics.

Several common types of error may interfere with the data collection when using a numerical rating instrument, as indicated by Ary, Jacobs, and Razavieh (1990). As positive responses may be seen as more socially acceptable, respondents may tend to give high ratings when they are in doubt, thus resulting in a generosity error. Conversely, respondents may be extremely critical on all issues, an error of severity. Or respondents, attempting to avoid being either overly critical or positive, may provide all neutral responses (the error of central tendency). For this study, these possible errors were combated by emphasizing in the cover letter the need for constructive criticism of the model. The respondents were instructed to consider each item separately. In addition, the instrument was designed to minimize the occurrence of response sets by providing two different scale formats (i.e., "too little" to "too much" for the quantity scale and "poor" to "excellent" for the quality scale).
Instrument 3: The third instrument was developed to measure the opinion of the panelist concerning the total performance of the guidebook. This instrument was a written, five item rating scale. The performance was rated on the following criteria: 1) objective, 2) feasible, 3) efficient, 4) flexible, and 5) useful. The mean response to these five categories was used to determine the individual's opinion of the performance. Relevant aspects of questionnaire design (e.g., space for additional comments, page layout, reading level, adequate directions) were considered.

Assessment of Validity and Reliability: Following the initial stage of designing the instruments, steps were taken to improve the face, content, and construct validity, as well as the reliability of the instruments. Five administrators of ROEP were asked to participate in the pilot test. Responses were gathered from four of the participants. A field study was conducted on five graduate students who have had experience with research methodology and residential outdoor education programs. The field and pilot tests also provided data for the initial assessment of the guidebook.

The pilot test was used to test for face and content validities. The pilot test respondents did not indicate that the content or face validities needed improvement. A pilot test was not conducted with a population of academic experts; however, the field test participants, who additionally commented on face validity, served to represent the academic panelists. The field test participants provided comments on the face, content, and construct validity of the instruments. Minor editorial changes and clarification of the directions were made to the questionnaire based on these comments. The comments from the field test did not indicate any problems with the content or construct validity of the instruments. Few of the pilot and field test respondents returned the guidebook and questionnaire by the deadline, indicating that panelists should be given more than two weeks to respond to the survey.
The *a priori* Cronbach's Alpha was .6 which is appropriate for this early stage of research (Nunnally, 1967). A test-retest was conducted with the pilot test participants to assist in making the instruments resistant to low reliability. However, since only three of the retests were returned, the reliability could not be determined prior to the study. The summated property of these instruments provides an additional means to inspect the reliability, that of internal consistency. After the data were collected, a Cronbach's Alpha was used to determine the internal consistency of each of the instruments and for each of the items in the semantic differential, Instrument Two. This analysis was conducted with the SPSS package. Items which did not meet the *a priori* criterion were not included in the analysis of the data. Data on the reliability of the instruments is reported in chapter four of this work. The Cronbach's Alpha was also used to bolster the instruments' construct validity.

**Conditions**

The instruments were delivered to the participants at their place of employment. For the field expert panelists, this was their resident outdoor education site. For the academic panelist, this was typically an office at a university. The guidebook was reviewed and the mailed questionnaires were completed under conditions established by the participants. The panelists were asked to return their responses on a specified date two weeks after they had received the guidebook and the questionnaire.
Data Analysis

The data were analyzed separately for each of the panels. For each panelist, the summated mean of each of the six semantic differential scales in Instrument One (vocabulary, cover, writing style, page layout, clipboard pages, and flow of the information) was calculated. Separate means were calculated for the summated scores on the quantity scale, the summated scores on the quality scale, each item in Instrument Two, and the summated scores on the performance scale.

Pearson product-moment correlation’s were calculated: 1) among the semantic differential items, 2) between the semantic differential items and the variables presentation, quality, quantity, and performance; 3) between each of the semantic differential items and each of the Instrument Two topics as rated by their quantity; 4) between each of the semantic differential items and each of the Instrument Two topics as rated by their quality; 5) between each of the items on the performance scale and topics on the content scale as rated by quantity; 6) between each of the items on the performance scale and topics on the content scale as rated by quality; and, 7) between each of the semantic differential items and each of the performance criterion. The \textit{a priori} alpha level was .05.

Responses to the open-ended comments on the questionnaire and in the guidebook were systematically coded and reported in response categories (i.e., suggested modifications to the presentation, questions, and concerns about the content, positive comments, etc.). Summaries, selected direct quotations, observations made by the researcher and trends found in the responses were recounted in the final report.
CHAPTER IV
RESULTS

The purpose of this study was to develop, evaluate, and refine a heuristic guide for the evaluation of residential outdoor education programs. This chapter presents the results of the final stage of the study, that of formal assessment of the guidebook by two panels of experts. The results presented herein answer research questions one and two.

Of fourteen potential academic panelists who agreed to review the guidebook, responses were collected from eleven persons. A majority of the respondents on the academic panel were professors in the field of outdoor education or recreation. The panel of field experts was composed of ten administrators who responded out of seventeen potential respondents. Of these administrators, six represented year-round programs and two represented seasonal programs; four represented programs with an educational emphasis and two represented programs with a recreational emphasis; three represented programs with a capacity for less than 200 persons, four represented programs with a capacity of from 200 to 500 persons, and one represented a large program with a capacity larger than 500. The first section of this chapter reports the results of the written questionnaire; the second section reports the findings from written comments on the guidebook or in accompanying letters; and, the third section reports the changes made to the draft as a result of the panelists' comments.
Findings From the Written Questionnaire

Written questionnaires were returned by eight members of the panel of field experts and by ten members of the panel of academic experts. A majority of the respondents provided responses to all of the items; however, in some cases, items or entire scales were left blank. As a result of the large amount of time required for the printing of the guidebook and shipping delays resulting from inclement weather, many of the panelists did not receive their packets as scheduled. One member of the academic panel and one of the field experts sent a message stating that they would not be able to return the materials by the deadline but that they would review the guidebook and return the materials as soon as possible. Several of the academic panelists returned only the questionnaire and indicated that they would return detailed comments on the draft at a later date. The response to one of the items provided by one academic panelist was unusable. In this case, the panelist's response to the useless-useful scale for the clipboard item was in direct opposition to an open-ended comment which was written below the scale. The results from the questionnaire are reported below as qualitative findings and quantitative findings.

**Qualitative Findings:** Open-ended remarks, written on the questionnaire, were categorized by the researcher. Typically these comments were general appraisals of an item such as "Good idea." In multiple cases, reviewers provided comments which directly opposed those of another reviewer. The following text details comments which were made about the: (1) vocabulary, (2) cover, (3) writing style, (4) page layout, (5) clipboard pages, (6) flow of information, (7) performance, and (8) additional comments.

The most common comment about the vocabulary, from the administrators and from several of the academic panelists was that the language was too academic. A typical remark was, "Extremely technical for the layperson [sic]." However, some of the
administrators suggested that the language was appropriate; one stated, "Even though the vocabulary suffers from typical academic cloudiness, it seems to be the best attempt I have seen at making the evaluation process seem workable. In contrast to the administrators' concerns about the technical nature of the wording, several of the academic panelists specifically stated that the vocabulary was appropriate for practitioners. Academic panelists also presented concerns about the uneven nature of the vocabulary, the need for definitions of terms when they are first mentioned, consistency of terminology, and the need for corrections in spelling.

A majority of the comments about the cover were negative. Two administrators disliked the acronym. One suggested: "That big on' ROPE acronym is obnoxious to me. I would emphasize the full title, then put (ROPE) in parenthesis or not use the acronym at all." One remarked that the cover mislead the reader by indicating that the guidebook would present simple instructions for evaluation. Negative remarks from the academic panelists included: "Too dark," and "The diagram should be labeled and explained." Four positive comments were written about the cover and the acronym.

The main concerns expressed by both panel of experts about the writing style were similar to the concerns about the vocabulary. Three reviewers suggested that the writing style was too academic, and three commented that it varied too much between different sections of the text. The academic panelists also noted: "Some tense changes were confusing," "There are many sentences ending in prepositions," and "Beware of sexist language." A majority of the positive remarks were from the academic panel and focused on the examples, for instance, one academic reviewer provided the remarks, "Beautiful, friendly metaphors and examples." Two of the administrators provided positive comments about the writing style; one wrote, "I was interested enough that I forgot to comment because I was reading for interest instead of reviewing."
Three types of criticisms of the page layout were provided: not user-friendly, graphics, and organization. One of the administrators felt that the layout was not user-friendly, one commented that the graphics did not always match the text, and one suggested that figures be separated from the text with a box. One of the academic panelists wrote that more graphics were needed, one commented that the second section, and the appendices could be intimidating to non-academicians, and one academic panelist suggested that better organization was needed in the techniques section. Three panelists' remarks indicated that the page layout was "o.k.," and one panelist commended the format, stating, "Very friendly format. I particularly appreciated the full references within the text AND at the end."

In general, comments about the clipboard pages were positive. However, two of the academic panelists and one of the administrators expressed that the clipboard pages would not be used. One panelist remarked that the clipboard header looked like a telephone. A typical positive remark was: "Nice approach to get the reader involved." A suggestion from an administrator was that these pages would be more useful for a workshop. This panelist also suggested that if the clipboard pages were clamped together and were detachable, they would be more useful.

Two respondents from each of the panels reported concern that the flow of the information was too difficult for novices. A typical remark was: "I do have a concern that it may be entirely too much for the average practitioner." An extreme comment was: "Overkill. The reader (this reader) was ready to scream with the plodding pace and repetition." Five panelists (two administrators and three from the academic panel) reported that they thought the flow of information was fine; however, several of these respondents qualified their responses, noting that they have had academic experience with either research or evaluation.
Three remarks from program administrators indicated that the performance of the guidebook did not meet their expectations as it presented more information than they wanted. The general emphasis of these comments were that the panelist was "overwhelmed." One administrator commented that the guidebook would be useful for designing an evaluation, and one member of the academic panel responded that it could be beneficial for students in research and evaluation courses, stating, "As a professor I thought it was thorough w/o [without] being overdone. Nice job!"

Many recommendations were provided in the "Additional comments" section. Five critical comments (four of which were from academicians) stated that the guidebook seemed like a textbook and did not provide "clear, concise, and do-able guidelines." Others from the academic panel commented on information which they felt should be included (such as guidelines for questionnaire construction) or sections which they felt should be either eliminated or expanded (such as the discussion of the Thurstone scale). One administrator suggested that section one was unnecessary and could be condensed and placed into an appendix. Four of the administrators indicated that the guidebook was educational, including the comment, "It helps make a complex subject very understandable." Two members of the academic panel remarked that the guidebook did not address methods for analyzing and reporting data. Three of the academic panelists provided generally positive statements such as "I like and appreciate this evaluation booklet."

In summary, the qualitative data provided on the questionnaire offered both general and specific criticisms and commendations. Recurring patterns which were found in these results were: 1) the vocabulary was too academic; 2) the cover was not attractive; 3) opinions about the clipboard pages were positive; and, 4) the guidebook might be more appropriate as a text for evaluation or research courses.
Quantitative Findings: The opinions and reasoned professional judgment of the panelists toward the presentation, the content, and the performance of the guidebook were quantified through the use of three instruments which were combined into one written questionnaire. The results for each panel of experts were analyzed separately. Descriptive statistics (means and correlations) were used for the analysis. Summarizations of the results are presented in Tables 1 through 8.

Responses from the panel of field experts: The mean of the field experts' responses toward the presentation of the material was 4.77 with a standard deviation of .73 on the seven point semantic differential scale. (See Table 1.) Means for individual topics within the scale were: 5.13 for the clipboard, 4.42 for the cover, 4.76 for the flow of the information, 5.16 for the page layout, 4.53 for the writing style, and 4.56 for the vocabulary. These means represent nine respondents except in the case of the vocabulary as one respondent did not provide responses for this scale. The means for each of these items was above a 4.0, indicating slightly positive attitudes toward the presentation of the model. The Cronbach's Alpha for all of these items were above .8.
Table 1. Field Experts' Responses to the Presentation of the Guidebook

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>MEAN</th>
<th>STANDARD DEVIATION</th>
<th>N</th>
<th>CRONBACH'S ALPHA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summated Score on the Presentation</td>
<td>4.77</td>
<td>.73</td>
<td>9</td>
<td>.91</td>
</tr>
<tr>
<td>Clipboard Pages</td>
<td>5.13</td>
<td>.69</td>
<td>9</td>
<td>.85</td>
</tr>
<tr>
<td>Cover</td>
<td>4.42</td>
<td>.86</td>
<td>9</td>
<td>.89</td>
</tr>
<tr>
<td>Flow of Information</td>
<td>4.76</td>
<td>.99</td>
<td>9</td>
<td>.96</td>
</tr>
<tr>
<td>Page Layout</td>
<td>5.16</td>
<td>1.02</td>
<td>9</td>
<td>.96</td>
</tr>
<tr>
<td>Writing Style</td>
<td>4.53</td>
<td>1.05</td>
<td>9</td>
<td>.92</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>4.56</td>
<td>.96</td>
<td>8</td>
<td>.86</td>
</tr>
</tbody>
</table>

Means were computed on the basis of a seven point semantic differential scale.
The field experts' reasoned professional judgment of the quantity of various aspects of the guidebook was assessed through a five point rating scale. "Too little" was coded a one, and "too much" was coded a five. The mean of the responses from eight administrators on the quality of these items was a 3.39 with a standard deviation of .58 (Cronbach's Alpha = .85). (See Table 2.) The quantity of every item, except for that of "pictures/graphics," was rated between a 3 and 4 with standard deviations ranging from .35 to 1.07. (See Table 2.) This was a minor indication that too large a quantity of information had been included. Only for the "pictures/graphics" item, with a mean of 2.63 and a standard deviation of .92, did these respondents indicate the need for more material. The sections with the highest means, indicating too much information, were the "Evaluation Audience," "Purpose," and "Evaluation Design."

As shown in Table 2, the means for the quality of these items ranged from 2.87 for the "Pictures/Graphics" to 4.13 for the "Examples." "Poor" quality was coded a one, neutral was coded a three, and "Excellent" quality was coded a five. The "Examples," "Further References," and "Resources" sections all rated positively when the standard deviation was accounted for. The administrators' responses indicated that the quality of both the "Pictures/Graphics" and the "Glossary" were poor; however, the standard deviations of these means overlapped the neutral score.
Table 2. Field Experts' Responses to the Content of the Guidebook

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>MEAN</th>
<th>STANDARD DEVIATION</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summated Score on the Quantity (Alpha = .85)</td>
<td>3.43</td>
<td>.55</td>
<td>8</td>
</tr>
<tr>
<td>Evaluation Audience</td>
<td>3.88</td>
<td>.83</td>
<td>8</td>
</tr>
<tr>
<td>Purpose</td>
<td>3.87</td>
<td>.83</td>
<td>8</td>
</tr>
<tr>
<td>Sources of Data</td>
<td>3.25</td>
<td>.89</td>
<td>8</td>
</tr>
<tr>
<td>Evaluation Audience</td>
<td>3.88</td>
<td>.99</td>
<td>8</td>
</tr>
<tr>
<td>Resources</td>
<td>3.00</td>
<td>1.07</td>
<td>8</td>
</tr>
<tr>
<td>Time Frame</td>
<td>3.75</td>
<td>.89</td>
<td>8</td>
</tr>
<tr>
<td>Techniques</td>
<td>3.50</td>
<td>1.07</td>
<td>8</td>
</tr>
<tr>
<td>Glossary</td>
<td>3.38</td>
<td>.92</td>
<td>8</td>
</tr>
<tr>
<td>Appendices</td>
<td>3.63</td>
<td>1.06</td>
<td>8</td>
</tr>
<tr>
<td>Further References</td>
<td>3.25</td>
<td>.71</td>
<td>8</td>
</tr>
<tr>
<td>Pictures/Graphics</td>
<td>2.63</td>
<td>.92</td>
<td>8</td>
</tr>
<tr>
<td>Examples</td>
<td>3.13</td>
<td>.35</td>
<td>8</td>
</tr>
</tbody>
</table>

Mean is based on a range of responses from (1) "Too Little" to (5) "Too Much."
Table 2. (Continued)

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>MEAN</th>
<th>STANDARD DEVIATION</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summated Score on the Quality (Alpha = .84)</td>
<td>3.38</td>
<td>.58</td>
<td>8</td>
</tr>
<tr>
<td>Evaluation Audience</td>
<td>3.62</td>
<td>1.06</td>
<td>8</td>
</tr>
<tr>
<td>Purpose</td>
<td>3.14</td>
<td>1.21</td>
<td>7</td>
</tr>
<tr>
<td>Sources of Data</td>
<td>3.29</td>
<td>1.38</td>
<td>7</td>
</tr>
<tr>
<td>Evaluation Design</td>
<td>3.25</td>
<td>1.28</td>
<td>8</td>
</tr>
<tr>
<td>Resources</td>
<td>3.75</td>
<td>.71</td>
<td>8</td>
</tr>
<tr>
<td>Time Frame</td>
<td>3.00</td>
<td>.93</td>
<td>8</td>
</tr>
<tr>
<td>Techniques</td>
<td>3.50</td>
<td>1.31</td>
<td>8</td>
</tr>
<tr>
<td>Glossary</td>
<td>2.88</td>
<td>.64</td>
<td>8</td>
</tr>
<tr>
<td>Appendices</td>
<td>3.38</td>
<td>.74</td>
<td>8</td>
</tr>
<tr>
<td>Further References</td>
<td>3.75</td>
<td>.71</td>
<td>8</td>
</tr>
<tr>
<td>Pictures/Graphics</td>
<td>2.87</td>
<td>.83</td>
<td>8</td>
</tr>
<tr>
<td>Examples</td>
<td>4.13</td>
<td>.83</td>
<td>8</td>
</tr>
</tbody>
</table>

Mean is based on a range of responses from (1) "Poor" to (5) "Excellent."
The panel of field experts rated the overall performance of the guidebook as 3.39 with a standard deviation of .92 on a five point scale. (See Table 3.) "Poor" was coded a one, and "Excellent" was coded a five. The administrators' questionnaire also included questions about the panelists' inclination to buy or to use the guidebook. Only one of the administrators would have bought the guidebook, while five reported that they would have used it if it was given to them.

### Table 3. Field Experts' Responses to the Performance of the Guidebook

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>MEAN</th>
<th>STANDARD DEVIATION</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summated Score on the Performance (Alpha = .95)</strong></td>
<td>3.39</td>
<td>.92</td>
<td>8</td>
</tr>
<tr>
<td>Objectivity</td>
<td>4.25</td>
<td>.89</td>
<td>8</td>
</tr>
<tr>
<td>Feasibility</td>
<td>2.88</td>
<td>1.13</td>
<td>8</td>
</tr>
<tr>
<td>Efficiency</td>
<td>2.88</td>
<td>.83</td>
<td>8</td>
</tr>
<tr>
<td>Flexibility</td>
<td>3.86</td>
<td>1.07</td>
<td>7</td>
</tr>
<tr>
<td>Usefulness</td>
<td>3.25</td>
<td>1.28</td>
<td>8</td>
</tr>
</tbody>
</table>

Mean is based on a range of responses from (1) "Poor" to (5) "Excellent."
In order to address research question number two, Pearson product-moment correlation coefficients were calculated to determine relationships among the variables. Significant correlations (Alpha = .05) are shown numerically in Table 4. In the discussion of these correlations, the impact of the small sample size should be recognized. Note, as a random sample was not selected, the findings for the significance of these correlations are not conclusive.

The responses from the field experts indicated that the overall quality was strongly correlated with the overall performance (r=.91) and the overall presentation (r=.86) of the guidebook. These panelists' perceptions of the overall quality of the guidebook also varied significantly with several of the individual items within the presentation instrument: the flow of the information (r=.71), the page layout (r=.85), the writing style (r=.91), and the vocabulary (r=.80). Additionally, the perceived quality was significantly positively correlated with the field panelists' perceptions of the feasibility (r=.81), the efficiency (r=.82), the flexibility (r=.88), and the usefulness of the guidebook (r=.91).

The administrators' perceptions of the performance of the guidebook were significantly correlated with their perceptions of both the presentation (r=.87) and the quality (r=.91) of the guidebook. Significant relationships were also found between the performance variable and each of the items on the presentation variable, except for the "cover" item. The perception of the performance varied with each of the following: clipboard (r=.71), flow of the information (r=.81), page layout (r=.84), writing style (r=.92), and vocabulary (r=.87).
Table 4. Correlation Coefficients For Data from the Panel of Field Experts

<table>
<thead>
<tr>
<th></th>
<th>Quality</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation</td>
<td>.86*</td>
<td>.87*</td>
</tr>
<tr>
<td>Clipboard</td>
<td>.66</td>
<td>.71*</td>
</tr>
<tr>
<td>Flow of Information</td>
<td>.71*</td>
<td>.81*</td>
</tr>
<tr>
<td>Cover</td>
<td>.49</td>
<td>.39</td>
</tr>
<tr>
<td>Page Layout</td>
<td>.85*</td>
<td>.84*</td>
</tr>
<tr>
<td>Writing Style</td>
<td>.91*</td>
<td>.92*</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.80*</td>
<td>.87*</td>
</tr>
<tr>
<td>Quality</td>
<td>1.00</td>
<td>.91*</td>
</tr>
<tr>
<td>Performance</td>
<td>.91*</td>
<td>1.00</td>
</tr>
<tr>
<td>R1</td>
<td>.50</td>
<td>.72*</td>
</tr>
<tr>
<td>R2</td>
<td>.81*</td>
<td>.85*</td>
</tr>
<tr>
<td>R3</td>
<td>.82*</td>
<td>.91*</td>
</tr>
<tr>
<td>R4</td>
<td>.88*</td>
<td>.90*</td>
</tr>
<tr>
<td>R5</td>
<td>.91*</td>
<td>.96*</td>
</tr>
</tbody>
</table>

**KEY**

Presentation = Overall presentation  
Quality = Overall quality  
Performance = Overall performance  
R1 = Objectivity of the guidebook  
R2 = Feasibility of the guidebook  
R3 = Efficiency of the guidebook  
R4 = Flexibility of the guidebook  
R5 = Usefulness of the guidebook

* = Significant at the .05 level
The results indicated that the perceptions of the overall performance did correlate significantly with each of the individual items within the correlation instrument. The correlations with the performance variable were: objectivity ($r = 0.72$), feasibility ($r = 0.85$), efficiency ($r = 0.91$), flexibility ($r = 0.90$), and usefulness ($r = 0.96$). The results for the "cover" item demonstrated no statistically significant correlations with either the quality or performance of the guidebook.

*Responses from the panel of academic experts*: The mean of the responses from ten of the academic experts to the presentation of the material was 5.12 on the seven semantic differential scale. (See Table 5.) For all items, these panelists responded more positively to the guidebook than did the field expert panelists. The means of the individual items were: 5.46 for the clipboard, 4.48 for the cover of the guidebook, 5.23 for the flow of the information, 5.41 for the page layout, 5.12 for the writing style, and 5.03 for the vocabulary. Accounting for the standard deviation, the responses for the clipboard, the page layout, and the overall presentation received scores above a 4.5, indicating positive opinions of these elements.
Table 5. Academic Experts' Responses to the Presentation of the Guidebook

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>MEAN</th>
<th>STANDARD DEVIATION</th>
<th>N</th>
<th>CRONBACH'S ALPHA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summated Score on the Presentation</td>
<td>5.12</td>
<td>.52</td>
<td>10</td>
<td>.75</td>
</tr>
<tr>
<td>Clipboard</td>
<td>5.46</td>
<td>.40</td>
<td>10</td>
<td>.69</td>
</tr>
<tr>
<td>Cover</td>
<td>4.48</td>
<td>.77</td>
<td>10</td>
<td>.69</td>
</tr>
<tr>
<td>Flow of Information</td>
<td>5.23</td>
<td>.77</td>
<td>10</td>
<td>.90</td>
</tr>
<tr>
<td>Page Layout</td>
<td>5.41</td>
<td>.83</td>
<td>10</td>
<td>.90</td>
</tr>
<tr>
<td>Writing Style</td>
<td>5.12</td>
<td>1.01</td>
<td>10</td>
<td>.95</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>5.03</td>
<td>.80</td>
<td>10</td>
<td>.86</td>
</tr>
</tbody>
</table>

Means were computed on the basis of a seven point semantic differential scale.
The mean of the responses of five academic panelists to the quantity of information was a 2.77. (See Table 6.) The internal consistency for this scale with these respondents was very low. Many of the items contributed to the low alpha (e.g., resources, pictures/graphics, and techniques). These items were not deleted. As a result of the nature of this scale (the individual items may have had widely varying quantities) and the small number of respondents, this measure for reliability may not be valid.

The mean scores of the academic panelists' reasoned professional judgment of the quantity of the individual topics are presented in Table 6. Nine of these panelists provided responses to the five item scale. For only one item, "Evaluation Design" with a mean score of 3.11, did the academic reviewers indicate that too large a quantity of information was provided; however, the standard deviation of this was large, .93, implying much disagreement among the respondents. The lowest score, a 2.33, was for "Pictures/Graphics;" this score verified the panelists' desire for more illustrative materials. The scores for the quantity of the other items overlapped the neutral score and, thus, did not provide any significant indications.

The academic panelists' mean scores of the quality of individual topics were all positive. Yet as a result of the large standard deviations, only the scores for "Evaluation Audience," 3.89 and "Pictures/Graphics," 3.78, demonstrated positive opinions. (See Table 6.)
Table 6. Responses of the Academic Experts to the Content of the Guidebook

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>MEAN</th>
<th>STANDARD DEVIATION</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summated Score on the Quantity (Alpha = .39)</td>
<td>2.77</td>
<td>.23</td>
<td>9</td>
</tr>
<tr>
<td>Evaluation Audience</td>
<td>3.00</td>
<td>.50</td>
<td>9</td>
</tr>
<tr>
<td>Purpose</td>
<td>2.89</td>
<td>.78</td>
<td>9</td>
</tr>
<tr>
<td>Sources of Data</td>
<td>2.67</td>
<td>.50</td>
<td>9</td>
</tr>
<tr>
<td>Evaluation Design</td>
<td>3.11</td>
<td>.93</td>
<td>9</td>
</tr>
<tr>
<td>Resources</td>
<td>2.78</td>
<td>.44</td>
<td>9</td>
</tr>
<tr>
<td>Time Frame</td>
<td>2.89</td>
<td>.6</td>
<td>9</td>
</tr>
<tr>
<td>Techniques</td>
<td>2.44</td>
<td>.73</td>
<td>9</td>
</tr>
<tr>
<td>Glossary</td>
<td>2.56</td>
<td>.73</td>
<td>9</td>
</tr>
<tr>
<td>Appendices</td>
<td>3.00</td>
<td>.71</td>
<td>9</td>
</tr>
<tr>
<td>Further References</td>
<td>2.89</td>
<td>.33</td>
<td>9</td>
</tr>
<tr>
<td>Pictures/Graphics</td>
<td>2.33</td>
<td>.71</td>
<td>9</td>
</tr>
<tr>
<td>Examples</td>
<td>2.67</td>
<td>.50</td>
<td>9</td>
</tr>
</tbody>
</table>

Means are based on scores ranging from (1) "Too Little" to (5) "Too Much."
Table 6. (Continued)

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>MEAN</th>
<th>STANDARD DEVIATION</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summated Score of the Quality</strong> (Alpha = .91)</td>
<td>3.65</td>
<td>.64</td>
<td>9</td>
</tr>
<tr>
<td>Evaluation Audience</td>
<td>3.89</td>
<td>.6</td>
<td>9</td>
</tr>
<tr>
<td>Purpose</td>
<td>3.67</td>
<td>.87</td>
<td>9</td>
</tr>
<tr>
<td>Sources of Data</td>
<td>3.67</td>
<td>.87</td>
<td>9</td>
</tr>
<tr>
<td>Evaluation Design</td>
<td>3.44</td>
<td>1.01</td>
<td>9</td>
</tr>
<tr>
<td>Resources</td>
<td>3.78</td>
<td>.83</td>
<td>9</td>
</tr>
<tr>
<td>Time Frame</td>
<td>3.56</td>
<td>.73</td>
<td>9</td>
</tr>
<tr>
<td>Techniques</td>
<td>3.56</td>
<td>1.24</td>
<td>9</td>
</tr>
<tr>
<td>Glossary</td>
<td>3.33</td>
<td>.87</td>
<td>9</td>
</tr>
<tr>
<td>Appendices</td>
<td>3.67</td>
<td>.87</td>
<td>9</td>
</tr>
<tr>
<td>Further References</td>
<td>3.67</td>
<td>.71</td>
<td>9</td>
</tr>
<tr>
<td>Pictures/Graphics</td>
<td>3.78</td>
<td>.67</td>
<td>9</td>
</tr>
<tr>
<td>Examples</td>
<td>3.78</td>
<td>1.30</td>
<td>9</td>
</tr>
</tbody>
</table>

Means are based on scores ranging from (1) "Poor" to (5) "Excellent."
Nine academic panelists provided responses to the instrument on the overall performance of the guidebook. The summated score for the performance of the guidebook was 3.67. (See Table 7). All of the items within this questionnaire were rated positively; yet, there were large standard deviations. The item "Objectivity" received a significant positive score of 4.11.

**Table 7. Academic Experts' Responses to the Performance of the Guidebook**

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>MEAN</th>
<th>STANDARD Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summated Score on the Performance</strong> (Alpha = .87)</td>
<td>3.67</td>
<td>.87</td>
<td>9</td>
</tr>
<tr>
<td>Objectivity</td>
<td>4.11</td>
<td>.78</td>
<td>9</td>
</tr>
<tr>
<td>Feasibility</td>
<td>3.11</td>
<td>1.27</td>
<td>9</td>
</tr>
<tr>
<td>Efficiency</td>
<td>3.78</td>
<td>.97</td>
<td>9</td>
</tr>
<tr>
<td>Flexibility</td>
<td>3.89</td>
<td>.78</td>
<td>9</td>
</tr>
<tr>
<td>Usefulness</td>
<td>3.44</td>
<td>1.42</td>
<td>9</td>
</tr>
</tbody>
</table>

Mean is based on a range of scores from (1) "Poor" to (5) "Excellent."
Correlations were computed for the academic panelists' perceptions of the overall quality and overall performance of the guidebook with their perceptions of: (1) the overall presentation, (2) the individual items in the presentation scale, (3) the overall performance, and (4) the individual items in the performance scale. (See Table 8.) Correlations which were significant at a .05 alpha level are reported herein. Again, it must be recognized that these correlations are based on the results from a small, non-random sample.

The academic panelists' perceptions of the overall quality of the guidebook significantly varied with their perceptions of both the presentation (r=.79) and the performance (r=.81) of the guidebook. The results indicated that the perceptions of the quality were related to those of the flow of the information (r=.91), the writing style (r=.89), and the vocabulary (r=.94). One significant correlation was found between the academic panelists' responses to the quality of the guidebook and to those responses to individual items within the performance instrument was with the flexibility of the guidebook (r=.73).

For the academic panelists, the responses to the performance were positively correlated with the responses to the quality but not with the responses to the presentation of the guidebook. Of the individual items in the presentation instrument, only the responses to the writing style and to the vocabulary demonstrated significant positive correlations with these panelists' perceptions of the quality. The overall performance was related to the feasibility (r=.96), the efficiency (r=.88), the flexibility (r=.67), and the usefulness (r=.90) of the guidebook. The responses to the items clipboard, cover, page layout, and objectivity of the guidebook did not indicate any significant correlations with either the quality or the performance of the guidebook.
Table 8. Correlation Coefficients For Data from the Panel of Academic Experts

<table>
<thead>
<tr>
<th></th>
<th>Quality</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation</td>
<td>0.79*</td>
<td>0.40</td>
</tr>
<tr>
<td>Clipboard</td>
<td>-0.09</td>
<td>0.24</td>
</tr>
<tr>
<td>Flow of Information</td>
<td>0.91*</td>
<td>0.55</td>
</tr>
<tr>
<td>Cover</td>
<td>-0.06</td>
<td>-0.30</td>
</tr>
<tr>
<td>Page Layout</td>
<td>0.21</td>
<td>-0.25</td>
</tr>
<tr>
<td>Writing Style</td>
<td>0.89*</td>
<td>0.70*</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>0.94*</td>
<td>0.79*</td>
</tr>
<tr>
<td>Quality</td>
<td>1.00</td>
<td>0.81*</td>
</tr>
<tr>
<td>Performance</td>
<td>0.81*</td>
<td>1.00</td>
</tr>
<tr>
<td>R1</td>
<td>0.69</td>
<td>0.61</td>
</tr>
<tr>
<td>R2</td>
<td>0.70</td>
<td>0.96*</td>
</tr>
<tr>
<td>R3</td>
<td>0.62</td>
<td>0.88*</td>
</tr>
<tr>
<td>R4</td>
<td>0.73*</td>
<td>0.67*</td>
</tr>
<tr>
<td>R5</td>
<td>0.61</td>
<td>0.90*</td>
</tr>
</tbody>
</table>

**KEY**

- Presentation = Overall presentation
- Quality = Overall quality
- Performance = Overall performance
- R1 = Objectivity of the guidebook
- R2 = Feasibility of the guidebook
- R3 = Efficiency of the guidebook
- R4 = Flexibility of the guidebook
- R5 = Usefulness of the guidebook

* = Significant at the .05 level
In summary, the quantitative results from the written questionnaire indicate slightly positive opinions of the guidebook. Specifically: 1) the academic experts' opinions of the performance, presentation, and quality variables were more positive than those of the field experts; 2) the cover received low ratings; 3) both panels indicated the need for more graphics; 4) the panels' responses differed in that the field experts reported that there was too much information, while the academic experts reported that there was too little information; and 5) several individual items were positively correlated with both the quality and the performance of the guidebook.
**Findings From Comments on the Guidebook**

Many reviewers provided comments directly on the guidebook. In general, the academic panel wrote more comments and provided remarks of a more in-depth nature about the content of the guidebook than did the field panelists. Members of both panels noted mechanical errors, suggested synonyms, and indicated sentences which were confusing. The following is a report of other types of comments from each of the panels.

**Panel of Field Administrators.** Although comments from the administrators about the content of the guidebook were generally limited, several substantial suggestions were made. (See Table 9.) Several administrators suggested additions to listings of the advantages and disadvantages of certain techniques. One recommendation was made to include "health and sanitation" to the list of topics relating to the evaluation of facilities. Additionally, it was noted that, in appendix three mathematical symbols were used with which the readers might not be familiar.

In numerous cases throughout the review process, panelists expressed opposing points of view. The remarks provided by three administrators on the inclusion of sociometric techniques provides an example. Two of these administrators volunteered that this technique could be useful, writing, "Good. Simple and relates to our camp." Another administrator expressed concern that this was not an appropriate technique to include, writing: "Why is this included in an analysis of an OE [outdoor education] program? It may be useful for grouping kids who come to a camp, although there are significant drawbacks to this."
Table 9. Field Experts' Comments Written on the Guidebook

<table>
<thead>
<tr>
<th>COMMENTS ON THE CONTENT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Suggested Additions:</strong></td>
</tr>
<tr>
<td>• Additional advantages and disadvantages for unobtrusive observations and scaled items.</td>
</tr>
<tr>
<td>• The topics &quot;health and sanitation&quot; for the evaluation of facilities and &quot;knowledge&quot; for the evaluation of staff.</td>
</tr>
<tr>
<td><strong>Questions and Concerns:</strong></td>
</tr>
<tr>
<td>• The mathematical symbols in the sample size section are unfamiliar.</td>
</tr>
<tr>
<td>• Some examples may not be appropriate for ROEP.</td>
</tr>
<tr>
<td>• The glossary is unnecessary.</td>
</tr>
<tr>
<td><strong>Positive Comments:</strong></td>
</tr>
<tr>
<td>• Figure 14, &quot;Standardized Indicators,&quot; is helpful.</td>
</tr>
<tr>
<td>• Appendix six, &quot;Hints for Successful Evaluations,&quot; is clear and helpful.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>COMMENTS ON THE PRESENTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Suggested Modifications:</strong></td>
</tr>
<tr>
<td>• Use more appropriate terminology (reviewers provided suggestions).</td>
</tr>
<tr>
<td>• Move references to the end of the document.</td>
</tr>
<tr>
<td>• Do not need to &quot;spell out&quot; definitions for &quot;quantitative&quot; and &quot;qualitative.&quot;</td>
</tr>
<tr>
<td>• Redraw Figure 12, &quot;Connecting Purposes to Techniques Through Sources of Data,&quot; which is confusing.</td>
</tr>
<tr>
<td>• Condense the reading material by eliminating redundant phrases.</td>
</tr>
<tr>
<td><strong>Positive Comments:</strong></td>
</tr>
<tr>
<td>• &quot;R.O.P.E. at a Glance&quot; provides a good summary.</td>
</tr>
<tr>
<td>• Analogies and examples are nice and clear.</td>
</tr>
<tr>
<td>• The pictures are good.</td>
</tr>
<tr>
<td>• Charts provide good summaries.</td>
</tr>
</tbody>
</table>
The most common recommendations from the field experts which pertained to the presentation were in the form of suggested terminology which the reviewer felt would be more appropriate. One of these panelists noted areas of the text which were repetitive and suggested that the redundant information be removed in order to reduce reading material. Two administrators indicated that Figure 12 "Connecting techniques to purpose through sources of data" was confusing, stating, "While these pages may make sense to you, they are almost overwhelming to a layperson."

A large majority of the administrators' positive statements guidebook addressed the examples, analogies, and pictures. These panelists identified the graphics which they found appealing and the sections of text which provided clear and helpful examples. The administrators also wrote that the charts, notably Figure 10 which summarized the advantages and disadvantages of the evaluation models and Figure 14 which listed useful standardized indicators, were especially helpful. Additionally, one administrator stated that "R.O.P.E. at a Glance" was a good summary, and one indicated that the icon on the first page of each chapter was appreciated.

One of the members of the panel of administrators provided comments in the form of a letter. This administrator did not review the guidebook, stating "it is far too academic for most practitioners in the field of camping. I know of very few directors who would actually read the manual, develop their evaluation technique and follow through to evaluate their programs." This respondent did feel that the guidebook would be helpful to researchers in the field. The respondent also found the illustrations clever, the layout attractive, and the cover inviting. This administrator suggested that a short pamphlet with samples of techniques and actual instruments would be more useful for practitioners.
Panel of Academic Experts: These reviewers provided extensive suggestions for rewording or rephrasing and noted mechanical errors. A majority of the non-mechanical comments which were provided by the academic panelists focused on additions to or revisions of the content of the guidebook. Tables 8 and 9 indicate the nature of these panelists' comments. A discussion follows of substantial suggestions or comments which were made by multiple reviewers.

The panel of academic experts suggested several substantial additions to the content of the guidebook. (See Table 10.) One reviewer recommended the addition of "administrative organization, policies, and procedures" as an object of evaluation for residential outdoor education programs. The inclusion of "tips and pointers" for the development of a behavior rating sheet and a list of opposing adjectives for use in a semantic differential scale was recommended by members of the panel. Several of these reviewers provided additional advantages or disadvantages which they felt should be included in the descriptions of the various evaluation techniques. One panelist noted the absence of information on the disadvantages of the CIPP model.

Typically, comments on a single aspect of the guidebook were made by only one reviewer. However, in five cases similar comments on the same topic were made by two reviewers. Two of the members of the academic panel expressed concerns about the written questionnaire section. One of these reviewers suggested including an emphasis on the complexity of these instruments, writing that inadequate written questionnaires were the result of "poorly conceptualized and designed questions by people who don't know what they're doing." Two panelists noted inaccuracies in the example which was provided for Guttman scaling. Two panelists indicated that the information on sample size would not be helpful for novices. And, two panelists questioned why addresses were not included for all of the standardized indicators.
<table>
<thead>
<tr>
<th>Suggested Additions:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Administrative organization, policies, and procedures to the objects of evaluation.</td>
</tr>
<tr>
<td>• &quot;Environmental harmony&quot; to issues of concern for an evaluation of the facilities.</td>
</tr>
<tr>
<td>• An explanation of the need to operationalize variables.</td>
</tr>
<tr>
<td>• An explanation of validity and reliability in the introduction.</td>
</tr>
<tr>
<td>• Additional characteristics to include in Figure 4, &quot;Comparison of formative and summative evaluations.&quot;</td>
</tr>
<tr>
<td>• Drawbacks in the information about the CIPP model.</td>
</tr>
<tr>
<td>• Information on appropriate response rates in appendix six, &quot;Hints for successful evaluations&quot;.</td>
</tr>
<tr>
<td>• Additional advantages and disadvantages for sections on: anecdotal records, written questionnaires, scaled items, Thurstone scales, semantic differentials, verbal techniques, focus groups, sociometric techniques, and standardized indicators.</td>
</tr>
<tr>
<td>• Additional references.</td>
</tr>
<tr>
<td>• Information about pre/posttest comparison group as a time frame possibility.</td>
</tr>
<tr>
<td>• More information on and references for content analysis.</td>
</tr>
<tr>
<td>• A section on ethical practices.</td>
</tr>
<tr>
<td>• A statement that participants should complete the instrument for a pilot test.</td>
</tr>
<tr>
<td>• Thirty-six suggestions for additional glossary entries and multiple suggestions for rephrasing existing definitions.</td>
</tr>
<tr>
<td>• References and clearer explanations for the sections on sociometric techniques and standardized indicators.</td>
</tr>
<tr>
<td>• Tips and pointers for the construction of a behavior rating sheet.</td>
</tr>
<tr>
<td>• A list of opposing adjectives which can be used in a semantic differential scale.</td>
</tr>
<tr>
<td>• Information on intervening factors such as weather and equipment.</td>
</tr>
</tbody>
</table>
Table 10. (Continued)

<table>
<thead>
<tr>
<th>Questions or Concerns:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Questions about the appropriateness of certain examples.</td>
</tr>
<tr>
<td>• The information about responsive evaluation does not contribute much.</td>
</tr>
<tr>
<td>• Do not recommend including questionnaires in newsletters.</td>
</tr>
<tr>
<td>• Chapter on &quot;Resources&quot; seems incomplete.</td>
</tr>
<tr>
<td>• Poor example on page 95. Visually impaired do not require special facilities.</td>
</tr>
<tr>
<td>• State precautions more strongly.</td>
</tr>
<tr>
<td>• Sociometric techniques are best if matched with observation.</td>
</tr>
<tr>
<td>• In appendix seven, &quot;Addresses,&quot; why are addresses not provided for all of the</td>
</tr>
<tr>
<td>standardized indicators?</td>
</tr>
<tr>
<td>• Some of the standardized indicators are &quot;dated&quot; and may be restricted to</td>
</tr>
<tr>
<td>certain geographical areas. Where can they be obtained? How much do they cost?</td>
</tr>
<tr>
<td>• Appendix 3 &quot;Sampling&quot; is more complex than other sections. The explanation</td>
</tr>
<tr>
<td>of sampling size would not be helpful to a novice. Expand this explanation.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Positive Comments:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The &quot;bull's eye&quot; model in the validity and reliability section is good.</td>
</tr>
<tr>
<td>• Good examples.</td>
</tr>
</tbody>
</table>
Five of the academic experts commented on the glossary for the purposes of either clarifying definitions or suggested additional entries. Several of the proposed entries were suggested by more than one reviewer.

The academic panelists also provided comments on the presentation of the guidebook. (See Table 11.) Several panelists offered terminology which would be more inclusive, alluding to the lack of examples of programs for adults or the over abundance of "camp" examples. Other remarks about the vocabulary included suggested language which was more precise and eliminated gender-specific pronouns. Two reviewers felt that there was too much white space on one page. Two of the reviewers also commented extensively on the lack of consistency in citation style. One of the members of the academic panel suggested that Figure 14, "Standardized Indicators," be moved to an appendix.

Two substantial positive comments were provided on the draft itself by the panel of academic experts. One panelist commented on the appropriateness of the use of bold and italics for emphasis; another wrote that the icon was appreciated.
Table 11. Academic Experts' Comments on the Presentation

<table>
<thead>
<tr>
<th>Suggested Additions:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Illustrations in the appendices.</td>
</tr>
<tr>
<td>• Inclusive terminology which applies to schools, camps, and programs for adults.</td>
</tr>
<tr>
<td>• A summary sentence in bold type for each evaluation model.</td>
</tr>
<tr>
<td>• More distinct visual cues in the techniques section to aid in organization.</td>
</tr>
<tr>
<td>• An &quot;other&quot; category with space for the user to add their own comments to the clipboards.</td>
</tr>
<tr>
<td>• Explanatory sentences to appendix seven, &quot;Addresses.&quot;</td>
</tr>
<tr>
<td>• An index.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Suggested Modifications:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Redraw the cover with the six steps of evaluation.</td>
</tr>
<tr>
<td>• Enlarge the icon.</td>
</tr>
<tr>
<td>• Correct the page numbers.</td>
</tr>
<tr>
<td>• Use more precise vocabulary.</td>
</tr>
<tr>
<td>• Be consistent with the citation style.</td>
</tr>
<tr>
<td>• Use the terms &quot;front-end,&quot; &quot;formative,&quot; &quot;summative,&quot; and &quot;remedial&quot; instead of &quot;pretest,&quot; &quot;posttest,&quot; and &quot;follow-up.&quot;</td>
</tr>
<tr>
<td>• Change language which could be construed as sexist.</td>
</tr>
<tr>
<td>• Distinguish between rating and ranking scales.</td>
</tr>
<tr>
<td>• Remove the story presented at the beginning of the focus group section.</td>
</tr>
<tr>
<td>• Remove white space on page 77.</td>
</tr>
<tr>
<td>• Do not use accreditation and certification as synonyms.</td>
</tr>
<tr>
<td>• Correct the example in the sociometric techniques section.</td>
</tr>
<tr>
<td>• Move Figure 14, &quot;Standardized Indicators,&quot; to an appendix.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Positive Comments:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Nice use of bold and italics for emphasis.</td>
</tr>
<tr>
<td>• Like the icon.</td>
</tr>
</tbody>
</table>
Four of the members of the academic panel included a letter in their response. These letters included both positive and negative comments, recommended changes, and suggested uses for the guidebook. One of the reviewers, whose expertise was in editing, detailed errors in the grammar. This respondent wrote that the "inexcusable intrusions of basic elements" would cause readers to doubt the credibility of the guidebook. Advice for improving the writing style was also included in the letter. The respondent did not complete the questionnaire, writing directly on the draft instead.

Two of the letters were very supportive of the guidebook, both indicating that the panelist was interested in using the guidebook in graduate or undergraduate level courses which they taught. These panelists also included recommendations such as including "more specific guidelines for interview protocols and data analysis/interpretation." One panelist sent a list of criteria for evaluating school-supported residential outdoor education programs which the panelist had developed and offered that the list could be included in the guidebook.

One final letter provided many insightful suggestions which could be incorporated in the text. Examples of these were: (1) "I would like you to explain the difference between an evaluation model and an evaluation technique;" (2) include a discussion of validity and reliability in the forest metaphor in the introduction, "Validity is whether your map and compass techniques will get you through the forest to where you want to go. Reliability is whether you can find your way every time in every forest...;" and, (3) "Can you add a small graphic next the name of each technique, to help distinguish these (for example a tape to indicate verbal, binoculars for behavioral, and [sic] feather pen for written, and a grab bag for misc. [miscellaneous techniques]?". The main concern which this reviewer expressed was that the "friendly voice" which was used in section one was "lost" in the second section.
In summary, the panelists provided ample comments on the guidebook to aid in the revision process. Generally: 1) the members of the academic panel provided significantly more comments and identified changes which they felt should be made to the content of the guidebook; 2) a majority of the field experts’ comments addressed the presentation of the guidebook; 3) in general, only one or two comments were provided on any area of concern; 4) for several of the areas of concern upon which multiple reviewers commented, contrasting remarks were reported; and, 5) many useful suggestions were provided in letters which were returned with the guidebook.

**Summary of Changes Made to the Draft**

A majority of the modifications of the guidebook, which resulted from the comprehensive review, were mechanical. Following is a description of other major modifications to either the content or the presentation of the guidebook.

**Content:** In order to clarify the purposes for and the limitations of the guidebook, a forward was added which stated that the guidebook was not designed to prepare the reader to develop well-tested evaluation instruments, rather, the guidebook was an introduction to the evaluation process. In the content section, major changes to the content were the addition of "Administration" as an evaluation object, "health and sanitation" and "environmental impact" to the aspects of a facilities evaluation, and "knowledge" to the aspects of a staff evaluation. In the evaluation design section, several characteristics were added to Figure 4, "Comparison of formative and summative evaluations," and information was added on the disadvantages of the CIPP decision making model. Modifications were made to several examples in order to make them more appropriate for the entire range of audiences for the guidebook.
In the techniques section, selected advantages and disadvantages were added to the descriptions of the following techniques: anecdotal records, written questionnaires, Thurstone scales, semantic differentials, verbal techniques, focus groups, sociometric techniques, standardized indicators, unobtrusive observations, and scaled items. The example of the Guttman scale was corrected. References were added to the descriptions of sociometric techniques, standardized indicators, and content analysis. In appendix 3, "Sampling," information detailing the determination of an adequate sample size was removed. A list of criteria for evaluating a school-related ROEP, offered by one of the panelists, was appended to the text. Forty-five additions were made to the glossary.

Presentation: The need for additional of graphics in the guidebook was indicated by both panels of reviewers. Therefore, pictures were added to pages (predominantly in the appendices) which lacked illustration. The cover of the guidebook was reorganized to emphasize the title, "Resident Outdoor Program Evaluation" instead of the Acronym, "R.O.P.E.". Additionally, a black header was added the pages of Section 2 and to the appendices to provide distinct visual cues. The font size was increased on the icons throughout the document. In numerous cases, the language was modified to be more inclusive, less repetitive, and more appropriate for the intended audience (for example the term "camper" was replaced with "participant").

Another modification of the presentation was the bolding of one summary sentence in the description of each of the evaluation models. On the clipboards, where an "Other" category was included, space was provided for the reader to write comments. Since Figure 12 was not understandable to all of the reviewers, five additional pages were included preceding the figure to convey the same information through more text, organized as sets and subsets. The citation style was rendered consistent throughout. Additionally, the text was rearranged to provide a more balanced layout, and the pagination was corrected.
CHAPTER V
CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS

This chapter includes a discussion of: 1) conclusions which were made about the R.O.P.E. model; 2) the development and research process which was used for this study; 3) recommendations for further modifications of the guidebook; and, 4) suggestions for future research on and developments for the evaluation of residential outdoor education programs.

Conclusions About the R.O.P.E. Model

Conclusions which can be drawn from this exploratory study are:

1) The panel of academic experts' opinions about the presentation and the performance of the guidebook were positive, and these panelists indicated that the quantity of information provided was slightly less than that which was needed.

2) The panel of field experts' opinions on the presentation and the performance of the guidebook were positive, and their responses indicated that the quantity of information provided was slightly more than necessary.

3) The reasoned professional judgment of the panelists was that the vocabulary and writing style were too technical and academic for the intended audiences and that a greater number of graphics should be added to the text.

96
4) There was disagreement among the panelists as to whether or not the guidebook was appropriate for the audience. Although there was a tendency for the academic panelists to applaud the text and for the administrators to criticize the level of technical material, even within these groups, there were opposing opinions.

5) In general, the reviewers indicated that the guidebook would be more appropriate as an introductory text for an evaluation or research course than for use by novice practitioners.

6) In response to the second research question, positive correlations were found among some of the variables. The results indicated that the panelists' perceptions of the quality of the guidebook varied with their perceptions of the overall presentation, the flow of the information, the writing style, the vocabulary, the overall performance, and the flexibility of the guidebook.

7) The panelists' perceptions of the overall performance of the guidebook positively correlated with those of the writing style, the vocabulary, the quality, the feasibility, the flexibility, and the usefulness of the guidebook.

Discussion of the Research Process

A five stage development and evaluation process was used for this study. Based on the literature (Hammerman & Chenery, 1985b; Henderson & Chenery, 1987), the need for a heuristic model for the evaluation of residential outdoor education programs was identified. In the initial stage of the research, specific literature which provided useful information proved to be elusive. As the field of educational evaluation is immense and as little meta-evaluation has been conducted concerning the evaluation design and techniques which are used with residential outdoor
education programs. this was a major task to identify, to assess the utility of, and to consolidate the information into a concise, usable form. Problems also arose from the lack of consistent terminology for, definitions of, or categorization of residential outdoor education programs. Thus, the literature of disciplines as diverse as social work and experiential education had to be searched.

The second stage of the research process entailed the drafting of the guidebook. The model which was developed and subsequently described in the guidebook was grounded in the literature about evaluation of resident outdoor education programs. The individual elements of the model (i.e., evaluation audience, purpose, design, etc.) parallel those of texts with similar purposes (e.g., Program Evaluation: A Practitioner's Guide for Trainers and Educators, by Brinkerhoff, Brethower, Hluchyj, and Nowakowski, 1983 and Evaluating Environmental Education in Schools. A Practical Guide for Teachers, by D.B. Bennett, 1984).

The R.O.P.E. model builds upon recently proposed models such as that of Chenery and Hammerman (1985a) and Henderson and Bialeschki (1993). Elements of the objectives-based model outlined by Chenery and Hammerman which were incorporated in the R.O.P.E. model are: an analysis of the programs' objectives, the establishment of evaluation questions, a consideration of the decisions to be made on the basis of the evaluation, inclusion of the actors involved in the evaluation, evaluation methods, and costs of the evaluation. As the R.O.P.E. model was created as an aid in the development of an evaluation design, and not as a comprehensive guide to the complete evaluation process, it did not include several of the elements presented by Chenery and Hammerman which addressed the use of the evaluation product.
In the second stage of the study, three of the four aspects of camp, staff, program, and facilities, which Henderson and Bialeschki (1993) listed in their evaluation matrix were included in the R.O.P.E. model as possible objects of an evaluation. The fourth aspect, that of the administration, was added to the model in the fourth stage of this research. These authors proposed that evaluation could be conducted through evaluation by objectives, outcomes, or standards. In the R.O.P.E. model, these issues were addressed as the evaluation object "specific program objective," the evaluation object "program," and the evaluation design "expert opinion," respectively. Thus, the framework for the R.O.P.E. model was developed from the existing foundation of evaluation theory.

The major challenge encountered in this stage of the study was the organization of the complex material. Initially, the intended format for the guidebook was that of a systematic decision tree, with an arrangement similar to that of a dichotomous key for the identification of an appropriate evaluation technique given certain resource constraints. However, the complexity of and multiple considerations associated with each of the evaluation techniques rendered this not feasible. This stage of the research resulted in a 103-page draft with nine chapters and limited page layout or graphics. The first eight chapters described elements which should be incorporated into the selection of an appropriate technique. The last chapter contained descriptions and listings of advantages and disadvantages of selected evaluation techniques. (Behavioral observations and written questionnaire question types were grouped together.)

The third stage of the research design was that of initial assessment of a draft of the R.O.P.E. model. This review of the guidebook proved to be quite helpful in not only determining changes to be made to the guidebook, but also in determining...
problems with the survey design. The greatest obstacle encountered in this stage of the research was in encouraging the reviewers to provide their comments promptly. The field and pilot test reviewers provided beneficial input on the need for better organization and for visual cues to the organization of the material. These respondents also enabled the researcher to determine the length of time which reviewers would need to be given.

The fourth stage of the research process, that of comprehensive review and revision of the guidebook, provided substantial information on the panelists' opinions and reasoned professional judgments of the guidebook. Through assessing the presentation, content, and performance of the guidebook, this review addressed the issues which were identified in the literature as appropriate for evaluating a guidebook or an evaluation model. Relevant issues which were incorporated were: scope, relevance, flexibility, feasibility, objectivity, representativeness, timeliness, accuracy, usefulness, appropriateness, design, and writing style.

The comprehensive review of the draft did not, however, provide unanimous recommendations for changes. Instead, the comments illustrated the differing needs of the various audiences. For example, those panelists who were professors (and who probably were primarily interested in research) indicated their preference for topics such as validity and reliability, while several administrators asked for more practical guidelines and instructions. The intent for the guidebook was to satisfy the needs of the diverse audiences through interpreting the "academic" topics, thus enabling readers from a variety of backgrounds to relate to the multiple aspects of the evaluation process.

In the final stage of the study, the draft of the guidebook was revised and a prepublication draft was produced. The information that was used in the review
The results from comments which were provided directly on the guidebook were the most easily applied in the revision process. In general, the other comments and the quantitative results were broad in scope; thus, while they provided significant insights into the nature of the guidebook, they were not as easily integrated. Such results might have been more useful if they had been collected earlier in the development process of the guidebook.

The major disadvantage of this emergent research design was the large time requirement. Several anecdotal reports from reviewers indicated that the review of the document was very time consuming (one reviewer mentioned working on the draft for three days). The nature of the guidebook placed a great demand on the respondents, thus lengthening the entire research process and reducing the likelihood of responses. Additionally, time delays in the printing and shipping of the guidebook further reduced the likelihood that panelists could respond by the stated deadline. The low response rates led to the gathering of information from only a selected number of individuals. The results from this study were not intended to be generalized beyond the respondents.

The use of both a panel of academic experts and a panel of field experts in the comprehensive review proved beneficial. As expected, the academic experts were able to provide valuable information about the content of the guidebook while the field experts were able to offer significant insights into the appropriateness of the presentation and the utility of the guidebook. It is interesting to note that the academic panelists were more likely than the administrators to comment that the guidebook was not appropriate for practitioners. One of the substantial results of
utilizing both the field expert and the academic expert populations for the comprehensive assessment was the suggestion for alternate uses of the guidebook (i.e., as a text for university courses). The inclusion of both of these audiences in the review process was most effective for obtaining a balance of recommendations for the heuristic model.

A number of administrators in this study indicated that did not feel qualified to conduct the types of systematic evaluations which were recommended in the guidebook. This reflects the overwhelming need for such a guidebook which can begin to educate and empower the administrators, thus fostering improved evaluations. Administrators of residential outdoor education programs typically are well-educated; Henderson and colleagues (1983) found that of two-thirds of the administrators whom they surveyed had earned at least a Master's degree. Therefore, it is felt that they can conduct or at least be meaningfully involved in the design of such evaluations (perhaps in partnership with research institutes); program administrators simply have not been provided adequate education to enable them to create successful evaluations. The guidebook is intended to provide administrators with a knowledge base for evaluation. All of the intricacies of the evaluation process can not be learned from a textbook, a workshop, or an academic course; rather they must be learned through experience. Administrators can best gain this experience through designing an evaluation, thus learning experientially. An educational guidebook can provide a foundation of knowledge about evaluation which will empower administrators to take this first step.
Recommendations for further modifications of the guidebook

The results of this study indicated that the guidebook may also be appropriate for use in an academic setting. Therefore, future modifications should insure that it is appropriate for such a use. As was indicated by the panelists, changes toward this end would include expanding the information on reliability and validity, content analysis, and sampling. Additionally, sections could be added on instrument development, extraneous variables, and data analysis.

Future revisions of the guidebook should include: 1) the inclusion of a discussion on ethical practices; 2) the selection of a more appealing color for the cover; 3) the incorporation of illustrative diagrams (notably in the discussion on sampling); 4) modifications of the vocabulary to make the language less academic; and, 5) the inclusion of a questionnaire to be used in a summative evaluation of the guidebook. Formal layout and design will be needed prior to the publication of this document.

A discussion on ethical practices could be appended to the guidebook in order to re-emphasize caveats which were mentioned throughout the guidebook, to suggest debriefing strategies, and to provide guidelines for the use of certain types of material (e.g., personal journals). Diagrams should be developed to illustrate difficult concepts or processes. For example, simple random, systematic random, and stratified random sampling, could be illustrated through the use of an example list of fictitious names. From this frame, names could be selected through each of these methods and presented as separate lists. The illustration would provide a comparison of these lists of names.

Throughout the guidebook, the terminology should be modified to be less academic. This can be accomplished through exchanging words which contain three
or more syllables for simpler synonyms and by eliminating or further interpreting technical terminology such as "triangulation."

A tear-out evaluation questionnaire should be appended to the text. This questionnaire could be used to collect data on readers' perceptions of the presentation, content, and performance of the guidebook. Readers of the guidebook could return the questionnaire to the authors anonymously, or they could provide their name and address if they were interested in participating in an in-depth summative evaluation of the guidebook.

**Suggestions for future research and developments**

**Suggestions for future research:** There exists a deficiency in research on evaluation of residential outdoor education programs. One of the primary needs within this field is for meta-evaluation of various techniques and models. Only limited research has been conducted on the development of evaluations for ROEP and little of this research provides information on the effectiveness of specific evaluation models or techniques. In order to improve residential outdoor education, effective evaluation strategies must be identified so that they can be adapted to individual programs.

**Suggestions for future developments:** Findings which addressed the second research question can provide guidance for authors of future texts. These findings indicated several elements which are related to the perceived quality and performance of a guidebook; the elements which strongly correlated to both the perceived quality and the perceived performance were: the vocabulary, the writing style, and the flexibility of the guidebook. It may be that a careful consideration of these elements during the developmental stages of educational guidebooks will lead to improved acceptance of the materials.
An outstanding need of administrators in the field of residential outdoor education is for education about evaluation. As Henderson and Bialeschki stated, "Evaluation can help us [ROEP] to be more effective and efficient if it is used properly" (1993, p. 35). As adequate education about evaluation has not been provided to administrators, much of the potential for program improvement through the use of evaluations has been lost. Information should be disseminated through workshops, journal articles, guidebooks, formal education, or other appropriate means in order to improve the state of evaluation of these programs. This education should emphasize the importance of producing information which is useful to decision makers. Providing administrators with information about various evaluation models and techniques and with the skills to use a variety of these techniques, will enable them to design and maintain an evaluation process which is tailored to the specific needs of their program.

The R.O.P.E. guidebook was designed to begin to meet these educational needs. The purpose of this guidebook is to aid program administrators in the selection of an appropriate evaluation model and technique for their program; it was not designed to provide guidelines for instrument development. Since many of the panelists indicated that guidelines for instrument development and data collection were of greater importance, it is recommended that separate guidebooks be developed for these purposes. Ideally a series of texts or booklets, each focusing on a successive stage of the evaluation process (i.e., instrument design, data analysis, etc.), could be developed. The R.O.P.E. guidebook would be the introductory text in such a series.

A computerized version of the R.O.P.E. guidebook could be created which would allow for efficient access to the material. If the content of the guidebook were incorporated in a software program, the user could proceed through the sequence of
decision steps and could enter specific information about their program for each of the decisions. Such a program would allow the user to either access or by-pass detailed information and could identify appropriate techniques based on the user's inputs.

The R.O.P.E. guidebook may also be applicable to programs other than residential outdoor education programs. This guidebook may be useful for: environmental education programs; day-use outdoor education programs; interpretive programs at parks, museums, or zoos; or, extension education programs.

In addition to the use of this guidebook for programs which are similar to residential outdoor education, guidebooks should be developed which are designed for evaluation within each of these fields. This means for educating administrators may be especially appropriate for persons in the fields which were drawn upon for the creation of the R.O.P.E. guidebook (i.e., environmental education and extension education) as well as for any non-residential or non-formal education programs (such as interpretive programs, or day-long outdoor education experiences). The R.O.P.E. guidebook provides one model for informing the reader of the decision steps required for selecting an appropriate evaluation technique. This decision-tree concept should be adapted, condensed, or expanded to meet the need in a variety of related fields for education about program evaluation.

In conclusion, this study has provided several contributions. Foremost, this research project resulted in the creation of a guidebook which is useful to both administrators of residential outdoor education programs and to students of outdoor education and related fields. The study illustrated a heuristic model for the decision making process incorporated in the selection of an evaluation technique. As a result of the indication that vocabulary, writing style, and flexibility are strongly related to the perceived quality and performance of the guidebook, this research provided
guidance for the development of similar guidebooks. Additionally, this research exemplified an emergent research process designed for the formative evaluation of the development of theoretically-grounded materials to be utilized by practitioners.
LIST OF REFERENCES


Gilbertson, K.L. Environmental literacy: Outdoor education training and its effect on knowledge and attitude toward the environment. Unpublished doctoral dissertation, The Ohio State University, Columbus, OH.


APPENDIX A

PANEL OF EXPERTS PARTICIPATING IN THE PILOT AND FIELD TESTS
PILOT TEST PARTICIPANTS:

Susan L. Johnson, National Wildlife Federation, Washington, DC

Frank H. Price, Jr., Pine Hill Adventures, Inc., Somerville, AL

William Wealand, Outdoor Ministries, United Church of Christ, Columbus, OH

Deborah S. Yandala, Cuyahoga Valley Association, Cuyahoga Valley National Recreation Area, OH

FIELD TEST PARTICIPANTS

Dean Freund, Director of Outdoor Education Department, Worthington School District, Columbus, OH

Laurel Hodory, Graduate Teaching Associate, School of Natural Resources, The Ohio State University

Claire Oberst, Graduate Research Associate, Department of Agriculture, The Ohio State University

Teresa Schretter, Graduate Research Associate, School of Natural Resources, The Ohio State University

Vera Volbrecht, Graduate Associate, School of Natural Resources, The Ohio State University
APPENDIX B

LETTER TO THE FIELD TEST PARTICIPANTS
Hi,

Thanks for agreeing to be a part of our field test! We are really excited about how this research is progressing. Let me catch you up on what we are doing. We have been writing a guidebook which is intended to help administrators of outdoor education programs (mostly camp directors) to conduct evaluations of their programs. Our research question is: How useful is this book? To do this, we will ask two groups to critique and to complete a brief survey on the guidebook. The populations will be: thirty "field experts," camp directors specifically chosen to represent a variety of programs (large, small, religious, recreational, etc.), and ten "academic experts," chosen from colleges and universities.

What we need your help with is the field test, a pretest to test the survey itself. Enclosed are the initial contact letter, the cover letter, a draft of the guide, and a copy of the questionnaire. Please review the guide and comment on the validity of the questionnaire. It will be helpful for you to complete the survey as the participants would. Most importantly, we need to know your opinions concerning whether or not the questionnaire measures what it is supposed to (content validity), if it is comprehensive (construct validity), and if it looks like it measures what it is supposed to (face validity). Any comments that you have about the guidebook, either about the content or the presentation would be greatly appreciated!

Knowing that the quarter will start to get more hectic in a couple of weeks, will you please return this entire packet by Tuesday October 26th? We need your responses before we can begin the actual survey. You can return the whole packet to Heather via the mailroom in 210 or to 386 Kottman, which ever is most convenient. For your efforts, we would like permission to include your name in the final publication, and we will send you a free copy.

THANKS SO MUCH!!!

Heather L. Martin
Graduate Research Associate

Gary W. Mullins, Ph.D.
Associate Professor
APPENDIX C

REQUEST LETTER TO THE ACADEMIC PANELISTS
December 10, 1993

Dear Dr.,

YOU have the opportunity to participate in an important study which will help the field of outdoor education. Let me explain. Because many residential outdoor education programs lack effective evaluation procedures, they often miss opportunities for program improvement. There is a need for guidelines for evaluation in this field. We are currently writing a guidebook (entitled Resident Outdoor Program Evaluation) which administrators of these programs can use to design an evaluation that will suit their needs.

We are asking a panel of experts to review the guidebook in order to ensure that it is both theoretically sound and applicable. Would you agree to be a part of this panel? We want to find out, from experts such as yourself, how to make the guidebook better meet the needs of the field of resident outdoor education.

By asking you to participate on this panel, we are asking you for an investment of your time and expertise. We estimate that reviewing the guide will take four hours. However, we are confident that the result, a guidebook that is tailored to the needs of residential outdoor educators, will be well worth the time and energy.

If you are able to be a part of the review panel, we request permission to list your name in the guidebook as a reviewer. We intend to publish the guidebook in Spring, 1994 and will send you a free copy of the final work.

Your expert review is needed. Please drop the enclosed postcard in the mail to let us know if you will take part in this study. We will mail you a draft copy of the guidebook and reviewing guidelines in January and will request its return after three weeks.

If you have further questions about the project, please give us a call at (614) 292-9828 or Fax (614) 252-7432.

Thank You!

Heather L. Martin  
Graduate Research Associate

Gary W. Mullins, Ph.D.  
Associate Professor
APPENDIX D

COVER LETTER TO THE ACADEMIC PANELISTS
January 11, 1994

Dear,

Thank you for taking the opportunity to be a part of this study! As we indicated in the last letter, we are preparing the Resident Outdoor Program Evaluation: A guidebook for evaluating outdoor education programs (ROPE Guidebook). The purpose of this guidebook is to help outdoor education programs (1) determine their evaluation needs and (2) match these with a useful evaluation technique(s) within their time and resource constraints.

We are asking you to read and to provide a critique of the draft guidebook. We are especially interested in your expert opinion of the content of the guidebook. Is the information on evaluation models and techniques appropriate, comprehensive, and relevant to the field? What evaluation models, techniques, or decision steps would you add or delete? Does the guidebook provide useful information? Please write your comments directly on the draft. Add your ideas and cross out paragraphs or sections that are not needed. Please comment freely! Your comments will be used to improve this book. We have also included a questionnaire which asks questions about specific parts or characteristics of the guidebook. Please complete this after you have reviewed the guidebook.

Please use the enclosed envelope to return only the pages of the draft upon which you have commented and the questionnaire. (Send us Xerox copies if you would like to keep a complete copy of the draft.) Please place it in the mail by Friday, February the 4th. Your feedback is needed to improve this document. Through sharing your constructive comments, you are building one more bridge on the road to better outdoor education programming.

Again, for your assistance, we would like to list your name in the guidebook and provide you with a copy of the ROPE guidebook, (which should be published in the Summer of 1994). After reviewing and returning the draft, return the enclosed postcard to grant us permission to list your name as a reviewer.

Thank you for your help!

Heather L. Martin
Graduate Research Associate

Gary W. Mullins, Ph.D.
Associate Professor
APPENDIX E

QUESTIONNAIRE DISTRIBUTED TO THE ACADEMIC PANELISTS
Thanks!

Please Return To:
Resident Outdoor Program Evaluation
The Ohio State University
School of Natural Resources
#2021 Coffey Road
#210 Kottman Hall
Columbus, OH 43210-1085
We are interested in your view of the presentation of the material for use by outdoor education practitioners. Consider how each of the following was presented: VOCABULARY, COVER, WRITING STYLE, PAGE LAYOUT, CLIPBOARD PAGES, and INFORMATION FLOW. Under each topic is a list of scales. Please rate the guide's presentation of that topic by placing an X in a space between the two adjectives. In the space below the ratings, please make any additional comments that you have about the topic.

Example:

**TYPESET**

APPROPRIATE :: X :: :: :: INAPPROPRIATE

TENSE :: :: :: :: :: X :: RELAXED

UNIMPORTANT :: X :: :: :: IMPORTANT

This person feels that the typeset is somewhat appropriate, very relaxed and moderately unimportant.

**VOCABULARY**

CLEAR :: :: :: :: :: UNCLEAR

APPROPRIATE :: :: :: :: :: INAPPROPRIATE

POOR :: :: :: :: :: RICH

PERTINENT :: :: :: :: :: NOT PERTINENT

AMBIGOUS :: :: :: :: :: PRECISE

TENSE :: :: :: :: :: RELAXED

GOOD :: :: :: :: :: BAD

DIFFICULT :: :: :: :: :: EASY

**ADDITIONAL COMMENTS:**

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________
<table>
<thead>
<tr>
<th>PAGE LAYOUT</th>
<th>ADDITIONAL COMMENTS:</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAOTIC</td>
<td>ORDERED</td>
</tr>
<tr>
<td>UNFRIENDLY</td>
<td>FRIENDLY</td>
</tr>
<tr>
<td>CROWDED</td>
<td>SPREAD OUT</td>
</tr>
<tr>
<td>AGREEABLE</td>
<td>DISAGREEABLE</td>
</tr>
<tr>
<td>CLEAR</td>
<td>UNCLEAR</td>
</tr>
<tr>
<td>GOOD</td>
<td>BAD</td>
</tr>
<tr>
<td>CONFUSING</td>
<td>UNDERSTANDABLE</td>
</tr>
<tr>
<td>DIFFICULT</td>
<td>EASY</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CLIPBOARD PAGES</th>
<th>ADDITIONAL COMMENTS:</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRUCTURED</td>
<td>UNSTRUCTURED</td>
</tr>
<tr>
<td>USELESS</td>
<td>USEFUL</td>
</tr>
<tr>
<td>FLEXIBLE</td>
<td>RIGID</td>
</tr>
<tr>
<td>APPROACHABLE</td>
<td>UNAPPROACHABLE</td>
</tr>
<tr>
<td>EASY TO USE</td>
<td>HARD TO USE</td>
</tr>
<tr>
<td>CONFINED</td>
<td>SPACIOUS</td>
</tr>
<tr>
<td>IRRELEVANT</td>
<td>RELEVANT</td>
</tr>
<tr>
<td>UNCLEAR</td>
<td>CLEAR</td>
</tr>
</tbody>
</table>
The following is a list of topics covered in the guide. Please rate each according to the quality and quantity of information provided. One being too little information or of poor quality; five being too much information or of excellent quality.

Be sure to consider the quality and quantity of each topic as separate from the rest.

<table>
<thead>
<tr>
<th>QUANTITY</th>
<th>TOPIC</th>
<th>QUALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Too little</td>
<td>Too much</td>
<td>Poor</td>
</tr>
</tbody>
</table>

(Circle your choice) (Circle your choice)

1 2 3 4 5 Evaluation Audience 1 2 3 4 5
1 2 3 4 5 Purpose 1 2 3 4 5
1 2 3 4 5 Sources of Data 1 2 3 4 5
1 2 3 4 5 Evaluation Design 1 2 3 4 5
1 2 3 4 5 Resources 1 2 3 4 5
1 2 3 4 5 Time Frame 1 2 3 4 5
1 2 3 4 5 Techniques 1 2 3 4 5
1 2 3 4 5 Glossary 1 2 3 4 5
1 2 3 4 5 Appendices 1 2 3 4 5
1 2 3 4 5 Further References 1 2 3 4 5
1 2 3 4 5 Pictures/Graphics 1 2 3 4 5
1 2 3 4 5 Examples 1 2 3 4 5

FLOW OF THE INFORMATION

VAGUE EASY TO USE STRONG UNPLEASANT ORDERED STAGNANT UNAMBIGUOUS SMOOTH
OBVIOUS HARD TO USE WEAK PLEASANT UNORDERED SWIFT AMBIGUOUS HALTED

ADDITIONAL COMMENTS:

______________________________________________________________________
______________________________________________________________________
______________________________________________________________________
______________________________________________________________________
______________________________________________________________________

10
Please rate the performance of the entire guide on the following criteria.

<table>
<thead>
<tr>
<th>CRITERION</th>
<th>PERFORMANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>Excellent</td>
</tr>
</tbody>
</table>

(Circle your choice)

- **OBJECTIVE**
  - Rating: 1
  - Ratings: 2, 3, 4, 5

- **FEASIBLE**
  - Rating: 1
  - Ratings: 2, 3, 4, 5

- **efficient**
  - Rating: 1
  - Ratings: 2, 3, 4, 5

- **Flexible**
  - Rating: 1
  - Ratings: 2, 3, 4, 5

- **Useful**
  - Rating: 1
  - Ratings: 2, 3, 4, 5

**ADDITIONAL COMMENTS:**

Use the space below to provide any further comments or concerns that you have about the guide. What other information do you feel should be included? What should be removed?

---

129
APPENDIX F

COVER LETTER TO THE FIELD EXPERT PANELISTS

Dear Director,

Thank you for taking the opportunity to be a part of this study! We are excited that you are interested in improving outdoor education through program evaluation. As we indicated in the last letter, we are preparing the Resident Outdoor Program Evaluation: A guidebook for evaluating outdoor education programs (ROPE Guidebook). The purpose of this guidebook is to help outdoor education programs (1) determine their evaluation needs and (2) match these with a useful evaluation technique(s) within their time and resource constraints.

What we are asking you to do is to read the enclosed draft, to use the worksheets to determine what evaluation techniques might be useful for your program, and to provide an overall critique of the guidebook. As you read the draft, please use the enclosed red pen to write your comments in the margins (feel free to edit). What parts do you like/dislike? Where is it difficult to understand? Does it provide useful information? Is it too long? Too technical? What parts were especially helpful? Please add your ideas and cross out paragraphs or even sections that you did not need. Please comment freely! Your comments will be used to improve this book. We have also included a questionnaire which asks questions about specific parts or characteristics of the guidebook. Please complete this after you have reviewed the guidebook.

Please use the enclosed envelope to return all of the pages of the draft that you have written on and the questionnaire. Please place it in the mail by Friday, February the 4th. Your feedback is needed to improve this document and to make it useful. Through sharing your constructive comments, you are building one more bridge on the road to better outdoor education programming.

Again, for your assistance, we would like to list your name in the guidebook and provide you with a copy of the ROPE guidebook, (which should be published in the Summer of 1994). After reviewing and returning the draft, just return the enclosed postcard to grant us permission to list your name as a reviewer.

Thank you for your help!

Heather L. Martin
Graduate Research Associate

Gary W. Mullins, Ph.D.
Associate Professor
APPENDIX G

QUESTIONNAIRE DISTRIBUTED TO THE FIELD EXPERT PANELISTS
Thanks!

What's Your Opinion

Please Return To:
Resident Outdoor Program Evaluation
The Ohio State University
School of Natural Resources
#9021 Coffey Road
#210 Kottman Hall
Columbus, OH 43210-1085
We are interested in your opinions on the way the material was presented. Consider how each of the following was presented: VOCABULARY, COVER, WRITING STYLE, PAGE LAYOUT, CLIPBOARD PAGES and INFORMATION FLOW. Under each topic is a list of scales. Please rate the guide's presentation of that topic by placing an X in a space between the two adjectives. In the space below the ratings, please make any additional comments that you have about the topic.

Example:

**TYPESET**

<table>
<thead>
<tr>
<th>APPROPRIATE</th>
<th>X</th>
<th>INAPPROPRIATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TENSE</td>
<td></td>
<td>RELAXED</td>
</tr>
<tr>
<td>UNIMPORTANT</td>
<td>X</td>
<td>IMPORTANT</td>
</tr>
</tbody>
</table>

This person feels that the typeset is somewhat appropriate, very relaxed and moderately unimportant.

**VOCABULARY**

<table>
<thead>
<tr>
<th>CLEAR</th>
<th>: : : : :</th>
<th>UNCLEAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPROPRIATE</td>
<td>: : : : :</td>
<td>INAPPROPRIATE</td>
</tr>
<tr>
<td>POOR</td>
<td>: : : : :</td>
<td>RICH</td>
</tr>
<tr>
<td>PERTINENT</td>
<td>: : : : :</td>
<td>NOT PERTINENT</td>
</tr>
<tr>
<td>AMBIGOUS</td>
<td>: : : : :</td>
<td>PRECISE</td>
</tr>
<tr>
<td>TENSE</td>
<td>: : : : :</td>
<td>RELAXED</td>
</tr>
<tr>
<td>GOOD</td>
<td>: : : : :</td>
<td>BAD</td>
</tr>
<tr>
<td>DIFFICULT</td>
<td>: : : : :</td>
<td>EASY</td>
</tr>
</tbody>
</table>

**ADDITIONAL COMMENTS:**

_________________________________________________________________

_________________________________________________________________

_________________________________________________________________

_________________________________________________________________

_________________________________________________________________
The following is a list of topics covered in the guide. Please rate each according to the quality and quantity of information provided. One being too little information or of poor quality; five being too much information or of excellent quality.

Be sure to consider the quality and quantity of each topic as separate from the rest.

<table>
<thead>
<tr>
<th>QUANTITY</th>
<th>TOPIC</th>
<th>QUALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Too little</td>
<td>Too much</td>
<td>Poor</td>
</tr>
</tbody>
</table>

(Circle your choice)

1 2 3 4 5 Evaluation Audience..... 1 2 3 4 5
1 2 3 4 5 Purpose.................1 2 3 4 5
1 2 3 4 5 Sources of Data........1 2 3 4 5
1 2 3 4 5 Evaluation Design.....1 2 3 4 5
1 2 3 4 5 Resources.............1 2 3 4 5
1 2 3 4 5 Time Frame.........1 2 3 4 5
1 2 3 4 5 Techniques..........1 2 3 4 5
1 2 3 4 5 Glossary.............1 2 3 4 5
1 2 3 4 5 Appendices...........1 2 3 4 5
1 2 3 4 5 Further References....1 2 3 4 5
1 2 3 4 5 Pictures/Graphics.....1 2 3 4 5
1 2 3 4 5 Examples............1 2 3 4 5

FLOW OF THE INFORMATION
VAGUE | OBVIOUS
EASY TO USE | HARD TO USE
STRONG | WEAK
UNPLEASANT | PLEASANT
ORDERED | UNORDERED
STAGNANT | SWIFT
UNAMBIGOUS | AMBIGUOUS
SMOOTH | HALTED

ADDITIONAL COMMENTS:

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
Please rate the entire guide on the following criteria.

<table>
<thead>
<tr>
<th>CRITERION</th>
<th>PERFORMANCE</th>
<th>Poor</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBJECTIVE</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>FEASIBLE</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>EFFICIENT</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>FLEXIBLE</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>USEFUL</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

ADDITIONAL COMMENTS:

Would you buy this guide? YES NO
If not, would you use it if it was given to you? YES NO

Please circle one adjective from each column below which best describes your program:

<table>
<thead>
<tr>
<th>Operating season</th>
<th>Emphasis</th>
<th>Capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seasonal</td>
<td>Recreational</td>
<td>Less than 200</td>
</tr>
<tr>
<td>Year-round</td>
<td>Educational</td>
<td>200 to 500</td>
</tr>
<tr>
<td>Therapeutic</td>
<td>Religious</td>
<td>More than 500</td>
</tr>
<tr>
<td>Religious</td>
<td>Other</td>
<td></td>
</tr>
</tbody>
</table>

Use the space below to provide any further comments that you have on the guide.

________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
APPENDIX H

FOLLOW-UP POSTCARD
Last week we mailed a packet to you containing the R.O.P.E. guidebook and a questionnaire. If you have not yet received it, or if you have any questions, please contact us at (614) 292-9828.

Thank-you!

_________________________  _______________________
Heather L. Martin            Gary W. Mullins, Ph.D.
Graduate Research Associate  Associate Professor
APPENDIX I

PANELISTS
Academic Expert Panelists

Dennis L. Elliott, Assistant Professor, The Ohio State University, Columbus, OH

Dr. Alan Ewert, USDA Forest Service, Washington, DC

Karla A. Henderson, University of North Carolina, Chapel Hill, NC

Marylin G. Hood, Ph.D., Hood & Associates, Columbus, OH

Phyllis Ford, Ph.D., Eugene, OR

Katherine James, Ph.D., California State, Long Beach, CA

Debra J. Jordan, University of Northern Iowa, Cedar Falls, IA

Clifford E. Knapp, University of Northern Illinois, Oregon, IL

Nancy Murphy, Ph.D., Antioch University, Seattle, WA

Paul F. Nowak, Ph.D., University of Michigan, Ann Arbor, MI

Deborah Simmons, University of Northern Illinois, Oregon, IL

Field Expert Panelists

Karen Currie Allen, C.C.D., Hopewell Camp and Conference Center, Oxford, MS

Robert L. Grabill, Pemigewassett, Hanover, NH

Joan Horn, Glen Helen Outdoor Education Center, Yellow Springs, OH

Linda Grier, Camp Friendship, Palmyra, VA

John P. Jarczyk, Algonquin, Algonquin, IL

Jerome J. Meeds, Catholic Youth Camps, Inc., McGregor, MN

Karl Olson, YMCA Camp Kitaki, South Bend, NE

Ronald Reed, Mohican School in the Out-of-Doors, Danville, OH

Roger Upcraft, Program Manager, Courage Camps, Golden Valley, MN
APPENDIX J

TIME LINE
ACTIVITY

STAGE ONE: REVIEW OF THE LITERATURE

- Establish need for a heuristic model
- Gather information on models and techniques
- Discuss evaluation with ROEP experts

STAGE TWO: DRAFT THE GUIDEBOOK

- Organize information and draft outline
- Complete initial draft
- Review initial draft

STAGE THREE: INITIAL ASSESSMENT

- Contact field and pilot test participants
- Mail field and pilot test
- Mail retests
- Reorganize, edit, develop page design

STAGE FOUR: COMPREHENSIVE ASSESSMENT

- Contact potential panelist
- Mail packets to panelists
- Mail follow-up postcard
- Analyze responses

STAGE FIVE: REVISION PROCESS

- Make revisions to the guidebook
- Make recommendations

DATES

- SPRING 1993
- March, 1993
- March - June, 1993
- April, 1993
- SUMMER 1993
- July-August, 1993
- September, 1993
- October, 1993
- FALL 1993
- October, 1993
- October 20, 1993
- November, 1994
- November-December, 1993
- WINTER 1993/1994
- December, 1993
- January 20, 1994
- January, 28, 1994
- February, 1994
- WINTER 1994
- February, 1994
- March, 1994
APPENDIX K

A GUIDEBOOK FOR EVALUATING RESIDENTIAL OUTDOOR EDUCATION PROGRAMS

(R.O.P.E.)

[Note: The dimensions of the original two-sided guidebook have been altered in order to meet the margin specifications of this document.]
Resident Outdoor Program Evaluation

(R. O. P. E.)

A guidebook for evaluating residential outdoor education programs
Resident Outdoor Program Evaluation

(R.O.P.E.)
A guide for designing evaluations for residential outdoor education programs

By Heather L. Martin, Gary W. Mullins, and Emmalou Norland

Partial support for the development of this document was provided by the Ohio Agricultural Research and Development Center, the School of Natural Resources, and The Ohio State University.
Panel of Reviewers

The following is a list of professionals in the fields of residential outdoor education or evaluation who graciously donated their time and expertise for a formal evaluation of this manuscript. The authors are indebted to these persons for their insightful contributions.

Karen Currie Allen, C.C.D., Hopewell Camp and Conference Center, Oxford, MS
Dennis L. Elliott, Assistant Professor, The Ohio State University, Columbus, OH
Dr. Alan Ewert, USDA Forest Service, Washington, DC
Dean Freund, Director of Outdoor Education, Worthington School District, Columbus, OH
Robert L. Grabill, Pemigewasset, Hanover, NH
Karla A. Henderson, University of North Carolina, Chapel Hill, NC
Laurel Hodory, Graduate Teaching Associate, School of Natural Resources, The Ohio State University, Columbus, OH
Marilyn G. Hood, Ph.D., Hood and Associates, Columbus, OH
Joan Horn, Director, Glen Helen Outdoor Education Center, Yellow Springs, OH
Katherine James, Ph.D., California State, Long Beach, CA
John P. Jareczk, Algonquin, Algonquin, IL
Susan L. Johnson, National Wildlife Federation, Vienna, VA
Debra J. Jordan, University of Northern Iowa, Cedar Falls, IA
Clifford E. Knapp, University of Northern Illinois, Oregon, IL
Jerome J. Meeds, Catholic Youth Camps, Inc., McGregor, MN
Paul F. Nowak, Ph.D., University of Michigan, Ann Arbor, MI
Claire Oberst, Graduate Research Associate, Department of Agriculture Education, The Ohio State University, Columbus, OH
Karl Olson, Outdoor Education Coordinator, YMCA Camp Kitaki, South Bend, NE
Ronald Reed, Mohican School in the Out-of-Doors, Danville, OH
Teresa Schretter, Graduate Research Associate, School of Natural Resources, The Ohio State University, Columbus, OH
Deborah Simmons, University of Northern Illinois, Oregon, IL
Roger Uperaft, Program Manager, Courage Camps, Golden Valley, MN
Vera Volbrecht, Graduate Research Associate, School of Natural Resources, The Ohio State University, Columbus, OH
William Wealand, Outdoor Ministries, United Church of Christ, Columbus, OH
Deborah S. Yandala, Cuyahoga Valley Association, Cuyahoga Valley National Recreation Area, OH
# Table of Contents

- **INTRODUCTION** ................................................................. xi
- **R.O.P.E. AT A GLANCE** ...................................................... xv
- **Section 1** ........................................................................... 17
  - AUDIENCE ........................................................................ 19
  - PURPOSE ......................................................................... 21
  - SOURCES OF DATA .......................................................... 31
  - DESIGN ............................................................................ 35
  - RESOURCES .................................................................. 53
  - TIME FRAME .................................................................. 55
- **Section 2** ........................................................................... 61
  - TECHNIQUES ................................................................... 63
    - BEHAVIOR OBSERVATIONS ............................................ 73
      - Anecdotal Records .................................................. 75
      - Direct Observation ............................................... 76
      - Games Observation ............................................. 78
      - Participant Observation ...................................... 80
      - Unobtrusive Observation .................................. 82
    - WRITTEN QUESTIONNAIRES ......................................... 84
      - Open-ended ............................................................. 86
      - Scaled Items ............................................................ 87
      - Likert and Likert-type .......................................... 88
      - Thurstone Scale .................................................... 91
      - Guttman Scale ........................................................ 92
      - Semantic Differential ........................................... 93
    - VERBAL TECHNIQUES .................................................... 95
      - Focus Group Interview ....................................... 97
      - Personal Interview ............................................. 99
    - MISCELLANEOUS TECHNIQUES .................................. 101
      - Artifact Analysis ................................................ 101
      - Check Sheet .......................................................... 104
      - Experience Sampling .......................................... 105
      - Outside Observer .................................................. 106
Q-Sorting .............................................................. 107
Self-Evaluation .................................................... 109
Sociometric Method ............................................. 110
Standardized Indicator ........................................ 111
Test ............................................................................. 117

Appendices ................................................................... 119

Appendix 1: Validity and Reliability .......................... 121
Appendix 2: Data Collection Times for Quasi-Experimental Designs .......................... 125
Appendix 3: Sampling ................................................. 129
Appendix 4: Non-Response Error ............................ 131
Appendix 5: Content Analysis ................................ 133
Appendix 6: Hints for Successful Evaluations ........... 135
Appendix 7: Evaluation Criteria for School Related Programs .......................... 137
Appendix 8: Addresses for the Publishers and Authors of Selected Standardized Indicators and Computer Software .......................... 139

GLOSSARY ................................................................. 141

REFERENCES .............................................................. 145
# Table of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIGURE 1</td>
<td>Six Steps in Evaluation</td>
<td>xii</td>
</tr>
<tr>
<td>FIGURE 2</td>
<td>Bennett’s Hierarchy of Evidence For Program Evaluation</td>
<td>24</td>
</tr>
<tr>
<td>FIGURE 3</td>
<td>Typical Sources of Data for Each Evaluation Purpose</td>
<td>32</td>
</tr>
<tr>
<td>FIGURE 4</td>
<td>Comparison of Formative and Summative Evaluations</td>
<td>36</td>
</tr>
<tr>
<td>FIGURE 5</td>
<td>Comparison of Fixed and Emergent Designs</td>
<td>37</td>
</tr>
<tr>
<td>FIGURE 6</td>
<td>Case Study Plan</td>
<td>39</td>
</tr>
<tr>
<td>FIGURE 7</td>
<td>Importance/Performance Results</td>
<td>44</td>
</tr>
<tr>
<td>FIGURE 8</td>
<td>Example of a Matrix for the Evaluation of a Canoeing Activity</td>
<td>45</td>
</tr>
<tr>
<td>FIGURE 9</td>
<td>Generalized Characteristics of the Evaluation Models</td>
<td>50</td>
</tr>
<tr>
<td>FIGURE 10</td>
<td>General Advantages and Disadvantages of the Evaluation Models</td>
<td>51</td>
</tr>
<tr>
<td>FIGURE 11</td>
<td>Models and When Data are Collected</td>
<td>56</td>
</tr>
<tr>
<td>FIGURE 12</td>
<td>Connecting Techniques to Purposes Through Sources of Data</td>
<td>69</td>
</tr>
<tr>
<td>FIGURE 13</td>
<td>Standardized Behavior Rating Forms</td>
<td>77</td>
</tr>
<tr>
<td>FIGURE 14</td>
<td>Standardized Indicators</td>
<td>112</td>
</tr>
<tr>
<td>FIGURE 15</td>
<td>Validity and Reliability</td>
<td>121</td>
</tr>
</tbody>
</table>
Forward

When addressing evaluation, administrators of residential outdoor education programs seem to be caught between a rock and a hard place. In order to encourage program improvement and to establish the worth of the program, evaluation must be carried out. However, the creation of a well-designed evaluation which can provide accurate information is a time-consuming task, requiring specific skills. Administrators may be tempted to find a well-tested evaluation design from another program and to apply it directly to their program. Yet, evaluations need to be tailored for each specific program and should evolve with the program and be adapted to the changing needs of decision makers. The challenge in the field of residential outdoor education is to create a workable yet sound evaluation process which will provide accurate and trustworthy information. Often times, such processes may require skills and expertise beyond those of the program staff.

This guidebook does not provide a "cure-all" evaluation design for residential outdoor education programs. Instead, it provides guidance for the first step toward the development of a successful evaluation. The guidebook will not prepare the reader to develop a well-researched evaluation instrument (however, references are provided which provide guidelines for the development process). Rather, this guidebook is designed to present some of the options for evaluation design from which program administrators may chose. Thus, the guidebook provides a foundation upon which administrators can formulate the research questions for building an evaluation design, specific to the needs of their program.
INTRODUCTION

EVALUATION

What is evaluation? Why is evaluation important? And, how does evaluation apply to me? These are common questions we have when we are told that we should be evaluating our program. Many of us shy away from evaluation. It is seen as too academic, and the misconception still exists that it can only be done by researchers with expensive equipment and considerable expertise. Equipment and knowledge are important aspects of an evaluation; yet, any person can conduct a successful evaluation of a residential outdoor education program as long as that person is committed to employing a careful scientific process and to using the outcomes. This may require some in-depth study or teaming with an evaluator/consultant.

Contrary to popular belief, evaluation is not a remote academic study; we do some form of it everyday. For example, if you were considering buying new tents for your program, you would conduct an informal evaluation. First, you look at the need for the equipment, and then, you look at the quality of the tents which are for sale. You decide whether a dome or a pup style would work best for your program’s trips. Finally, you make a decision about the value of the new tents relative to your budget. You’ve just completed the basic steps of an evaluation. But, when you think about evaluating your resident outdoor education program, you are not sure how to evaluate on such a large scale. This guidebook will lead you step-by-step through the planning of your Residential Outdoor Program Evaluation (ROPE). Some techniques and strategies are relatively straightforward; others are more complex and may require assistance from persons skilled in the evaluation process.

Who should read this:

The guidebook is designed for administrators and staff of resident outdoor education programs (or camps) who want to create, expand or improve their evaluation efforts. Students studying outdoor education may also find this material useful. ROPE applies to any program that supports formal or non-formal education “in, about, or for the out-of-doors,” the classic definition provided by Donaldson & Donaldson (1958, p. 63). Some examples of these programs are: school environmental education camps, organization-sponsored camps (i.e., 4-H, Boy Scouts, Girl Scouts, etc.), religious camps and retreats, private camps, adult adventure vacations, therapeutic camps and recreational summer camps. Although each of these programs may have unique goals and interests, all share their residential element and a belief in education through direct experience.
About Evaluation:

Evaluation is the combination of science and art to describe and understand something. Program evaluation is a specific type of evaluation. Ideally, program evaluation is an on-going, evolving process which looks at all aspects of a program. Program evaluations are designed to improve the program or to help in decision making. These evaluations can provide information on the inputs (what goes into the program). They can also provide us with process data. Process data is information about what actually is happening in the program. By gathering process data, we can discover both the planned and the unplanned effects of the program (i.e., a litter clean-up contest may be intended to increase the participants' environmental awareness, but it may also create a competitive atmosphere). Finally, program evaluation can indicate the short- and long-term results or outputs of our efforts (i.e., the actual changes in self-esteem, the improvements in cooperation, or the increase in knowledge). The ROPE Guidebook is a starting point to guide you through the planning of your evaluation.

Evaluation of educational programs has long been accepted as a means of providing information for program improvement and for determining the worth or value of the program for subsequent decision making (i.e., providing credibility). Through evaluation, you can (1) gain insight into positive and negative effects on the participants, (2) identify successful techniques and problem areas, and (3) document your effectiveness and efficiency for marketing and accountability purposes. Some program administrators have found that they can achieve higher goals and improve their participant return rate through using evaluations (Cheeney & Akers, 1987). Evaluation can be a powerful tool if it generates information that can and will be used.

**FIGURE 1. Six Steps in Evaluation**

An evaluation, like an outdoor education program, is dependent on many separate parts which must be linked together to form the whole. (See Figure 1 for six steps found in most evaluations.) No one would expect to have a successful program session without extensive planning, careful choice of staff, and modifications during the session; an evaluation requires the same efforts.

A successful evaluation is built on a sound design (Step 1) which provides a firm foundation for the entire process. The remainder of the evaluation (data collection, analysis, and report of the results) is then constructed upon that foundation. Finally, you use the results of the evaluation to make improvements or decisions about the program, you evaluate your evaluation (called meta-evaluation), and you begin planning a new evaluation.
This guidebook will address in detail the first step of the evaluation process, that of developing a design. The design is the blueprint for the entire evaluation and gives the evaluation focus and direction. Part of designing an evaluation is choosing an evaluation technique or techniques (i.e., methods for collecting information, such as interviews, questionnaires, or written tests) which can be used to collect the information that you need. This guidebook will describe a variety of techniques. To conduct a successful evaluation, you need to carefully choose a technique(s) that meets the needs of your program.

For example, new staff often have a hard time dealing with discipline problems because they typically rely on one method (i.e., they may take away swim time with little regard for the nature of the problem). This might work for ten year olds who are teasing each other, but it probably would not have the same effect on a teenage girl who is refusing to eat. An experienced staff member, on the other hand, knows many techniques to encourage the behavior that is wanted and will choose a method for dealing with each problem based on the individual and the particular situation. If an evaluator tries to use only one evaluation technique for every evaluation, there will be problems similar to those of the inexperienced staff member. An effective evaluator, like an experienced staff member, takes an in-depth look at the program in order to select an appropriate technique. Our goal is to help you to build successful evaluations through carefully selecting evaluation techniques which meet the needs of your program.

Think of this evaluation planning process as a hiking trip. You want to get from point A to point B, through a forest. The trail on which you start guides you until you come to an intersection. At that point, you must decide which new trail to take. You must make the right
choice at several intersections in order to reach point B. With an evaluation, you start with what
you already know about your program (point A), and sift through much information (the forest) to
a wealth of pertinent information (point B). The ROPE Guidebook is the trail system. At the
decision points (intersections), you must provide specific information about your program to help
you select the right path (an appropriate technique). Of course, there is no absolutely "right"
technique, but thoughtful choices will lead you through the forest of possible information to that
which is the most helpful.

The first section of the guidebook provides you with brief chapters that address each of
the main decision points. These include identifying who will use the results, the purposes for
your evaluation, where you can get the necessary information, how you will structure your
evaluation, and how you can make the best use of your resources. It is crucial that you think
through each of these decisions and make use of the worksheets provided. The worksheets are
headed with a picture of a clipboard (for an example, see page 20). [Note: If you wish to
maintain an unmarked master copy, the original worksheets may be copied for use.] The
worksheets will help you to identify the more effective technique(s) for your program. The second
section of the guidebook describes a number of specific techniques that can be used to gather
information. At the end of the guidebook you will find appendices (with specific information on
data analysis, sampling, etc.), a glossary, and further references. Throughout the guidebook,
additional references are suggested. A caveat with any guidebook such as this is that it is not a
panacea. This guidebook can not provide all of the knowledge necessary to be come an expert
evaluator. Users are encouraged to go to these references for more examples or in-depth
discussions of specific topics. References in bold are particularly helpful. Additionally, key
concepts are bolded in the text and are set apart in a box at the end of each chapter.
R.O.P.E. AT A GLANCE

Resident Outdoor Program Evaluation (ROPE) is a decision tree process, outlined in this guidebook, to promote the evaluation of resident outdoor programs. Evaluation can be a powerful tool for improving and making crucial decisions about your outdoor education program. The guidebook will help you plan your own evaluation, tailored to the specific needs of your program. The following sections are included:

SECTION 1: This section describes the decision making steps required for designing an evaluation.

AUDIENCE: Who will be using the information gathered through the evaluation? Early in your planning identify those persons who will use the evaluation results to make decisions about and to make improvements to the program (for example: staff, parents, and board of directors). This will help you concentrate on only the information that is needed.

PURPOSE: Why am I evaluating? Decide exactly what you want to know so your evaluation will be focused and efficient.

SOURCES OF DATA: Where can I get the best information? Gather information from as many relevant sources as possible (different groups of people, records, etc.), using your evaluation purpose(s) as a guide.

DESIGN: How do I create a framework for my evaluation? Consider your audiences' and your own intentions for the evaluation, approaches to planning, and preferences for a particular evaluation model.

RESOURCES: How can I do a thorough evaluation with limited resources (time, money, and evaluation expertise)? Through careful planning and selection of methods that take advantage of accessible resources, you can design a successful evaluation.

TIME FRAME: When, in the life of the program, should I actually do the evaluation? Consider four factors: your evaluation purpose, the evaluation model that you are working with, the availability of the staff who will conduct the evaluation, and when you have access to your sources of information.

SECTION 2: This section of the guidebook describes sets of specific techniques that can be used to evaluate resident outdoor education programs.

TECHNIQUES: There are many ways to gather information (e.g., questionnaires, interviews, observations, etc.). Which one(s) should I use? Descriptions, advantages, and disadvantages of many methods are provided in this guidebook. Compare these to the needs of your audience, your purpose(s), your design, the available resources, your evaluation teams capabilities, and your time frame to choose the one that best suits your specific needs.
Section 1
Decision Steps For Designing an Evaluation
To identify your evaluation audience, consider:

1. all of the program's stakeholders, and
2. those stakeholders with the greatest need for information.

First, consider who will actually use the evaluation. These persons are called the *evaluation audience*, and they are the ones who will make decisions about or improvements in the program based on the evaluation results. The evaluation must focus on answering their questions.

**WHO THEY ARE**

Any person who has an interest in the program or the evaluation of the program is considered a *stakeholder* and could be included in the evaluation audience (Worthen & Sanders, 1987). Stakeholders often include: participants, program staff and administrators, parents, teachers, school boards, camp boards, sponsors of the program, potential clients, potential staff, and the community. The evaluation audience consists of those stakeholders who are in the position to make decisions about the program and is often represented by the staff, board of directors, school boards, and parents.

**HOW TO CHOOSE THEM**

It would be great if an evaluation could answer all questions for all stakeholders. For example, both school boards and teachers may have questions about the effects of a particular resident outdoor education program on the students, and you would like to reassure both groups of the positive effects. However, these groups may place different emphasis on questions such as: "How much did students learn about geology?" versus "How nutritious were the meals?". Usually, this leads to a problem: it is not practical to try to answer everyone's questions. To use resources wisely, as you plan, focus on the questions of the key evaluation audience(s). Use the following worksheets to determine who these people are by identifying and ranking the stakeholders and their needs. The questions of those stakeholders who have the greatest need for information should guide the evaluation. Less urgent questions should be addressed as the resources permit.

*Design your evaluation to answer the questions of the people who will use the evaluation to make decisions about the program.*
STAKEHOLDERS

List all the people or groups of people who have an interest in your program (i.e., stakeholders). Be specific. Include sponsors, participants, and staff.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

POSSIBLE EVALUATION AUDIENCES

1) Place a check mark (✓) by the people on your list who might use the evaluation results when making decisions about or improvements to the program.

2) Interview staff, colleagues, and outsiders to identify decision makers, special interest groups, program advocates, and critics.

3) Observe the program and review organization documents (i.e., job descriptions) to identify formal and informal leaders, authorized decision makers, and actual decision makers.

4) Add those people identified in steps two and three to your list. Repeat step #1 for these newly identified stakeholders.

5) Rank order the stakeholders who were checked. Use one or more of the following methods:
   • By SPONSORSHIP: Whoever funds the evaluation or the program may deserve a higher rank.
   • By ABILITY TO USE: Persons who are in a decision-making position and who are willing to use the information should have the higher ranking.
   • By DEGREE OF AFFECT: The people most directly affected by the evaluation have the right to the higher rank (Brinkerhoff, Brethower, Hluchyj, & Nowakowski, 1983).

6) Select the audience(s) to whom you have given the highest priority as the focus for your evaluation. As you consider other factors, such as resources, you may alter these later.
In developing the purpose for your evaluation, consider

1. the intended use of the information produced,
2. your audiences needs, and
3. the object of your evaluation.

WHY AM I EVALUATING?

There are an infinite number of reasons to evaluate. In the introductory section, two of the uses of information gathered from residential outdoor education program evaluations were listed as: 1) program improvement and 2) determining the value of the program for decision making. More specifically, evaluators are often interested in determining the positive and negative as well as the unintended effects on different people associated with the program. The ideas provided in this guidebook may give you some guidance, but you need to identify a specific purpose for your evaluation. A clearly stated and focused purpose will be the backbone of your evaluation, giving structure and support with flexibility. This chapter will share with you examples of common ROPE purposes and will help you to identify a distinctive purpose for your evaluation.

Evaluations are conducted to answer the most important concerns of the evaluation audience(s). As an evaluator, your job is to identify the questions that the audience may have (e.g., Did the participants enjoy swim time? or Do they have fun playing group games?) and to combine them into a statement of purpose (i.e., to determine the participants' attitudes toward the program activities). In order to do this you must first gather information about the
audience's needs. This might be done formally, as through a questionnaire to school boards asking what documentation they are seeking, or informally, as through a conversation with parents to discover their questions concerning the programs' effects on their children. It is most important that you identify what your audience expects from the evaluation so that you answer not only your questions but those of the key stakeholders as well.

OBJECTS OF EVALUATION

One way to organize the purposes of evaluation is through the objects on which the evaluation focuses. The evaluation object is the concept, event or idea that is being evaluated. Most ROPEs focus on one of the following six objects: (1) the actual program, (2) program purpose, (3) facilities, (4) staff, (5) administration, and 6) one or several specific program objectives. Below are descriptions of these objects. Expand on and use these to organize the concerns of your audience(s) into a purpose.

1) Program:

If the object of the evaluation is the program, the evaluator asks the question: What is actually happening in this program? The evaluator's goal is to form an in-depth description of the program. This is most often accomplished through detailed monitoring, record keeping, or observation. The description that is generated should be tailored to the needs of the audience and may vary from a numerical description of attendance, accident reports or budget records to a completely qualitative passage reporting personal feelings and observations. The program description may be focused on the actual incidents and events at camp, the behavior of the participants, and/or the attitudes and thoughts of the stakeholders. Evaluations of the program as a whole are also an excellent way to determine the unanticipated effects of the program.

2) Program Purpose:

Some evaluations are designed specifically to examine the worth or value of the program's purpose and/or broad goals. This type of evaluation often involves difficult questions such as: which has higher priority, promoting group cooperation, or encouraging personal growth (hopefully these are mutually achievable). For gathering information about the program purpose, evaluators typically do not rely on input about the actual program. Rather, the evaluator often asks stakeholders questions such as: Should we emphasize behavior or attitude change in our environmental education program? Evaluations which focus on the program's purpose rely on criteria upon which the purposes are judged. These criteria may be created externally (e.g., through parents or experts in camping and education) or internally (e.g., by staff or board of directors).
3) Facilities:

The program's facilities are often the object of evaluations conducted by an outside agency such as the local health department or the American Camping Association (ACA). In order to assess the facilities, evaluators usually address whether or not the camp's physical facilities are meeting the needs of the participants. Common topics are user satisfaction, cleanliness, safety, appropriateness for program purposes and activities, environmental impact, and accessibility.

4) Staff:

Since the staff are responsible for bringing the program to life, they can be the object of a ROPE. A staff evaluation might focus on the performance or competencies of individuals or of the entire staff. For example, administrators may want to know if the summer staff are carrying out their duties and if not, why not. Or, a program director may want to determine how well individual counselors are communicating with campers so that the administration can provide additional training as needed. Skills which may be assessed include communications, technical, interpersonal, or conceptual (i.e., embracing the vision and traditions of the agency).

5) Administration:

An evaluation may concentrate on the administration and the organizational structure of a program. Relevant evaluation questions might be: (1) is the organizational structure effective and efficient? or, (2) Are the administrative policies and procedures pro environment, are they inclusive of all persons, and do they meet local, state, and federal guidelines and laws (i.e., ADA). For example, the board of directors of an outdoor school which has fifteen instructors and one administrator may be interested in determining whether their current organizational structure is meeting the needs of the staff.

6) Program Objectives:

Many times one or a set of the program's objectives is the object of the evaluation. Evaluations which focus on program objectives may determine: the extent to which the objectives have been achieved, why they have or have not been achieved, and/or how to better achieve them. For example, if a program's philosophy emphasizes increasing the participants' self-esteem, it is meaningful to stakeholders to know whether or not this is actually occurring. Typically, program objectives fall into one of these three categories: affective change (feelings, attitudes, or opinions), behavioral change (new or modified behaviors or skills) or cognitive change (analytical skills, knowledge, or problem-solving ability).
LEVELS OF ASSESSMENT

Determining a program's progress toward a purpose, goal, or objective can be confusing. One way to conceptualize a program's progress is as steps toward an ultimate end result. C.F. Bennett created a seven step hierarchy which illustrates stages of progress toward an end result. (See Figure 2.) This hierarchy has been used extensively by the Extension Service as a framework for program evaluation.

FIGURE 2. Bennett's Hierarchy of Evidence For Program Evaluation.


Let's look at how an imaginary camp, Camp Ladden, used these stages to assess its progress toward their objective of increasing participation in and enjoyment of lifelong outdoor leisure activities (i.e., canoeing, outdoor cooking, hiking, wildlife observation, etc.).

The Camp Ladden evaluators tabulated inputs such as staff, equipment, allotted program time, etc. which were required to plan and produce the program activities that promoted lifelong leisure activities.

Next, the evaluators reviewed all of the staff's activities, recording the activities and events like planning, marketing, and staff training, which supported the objective of increasing participation in and enjoyment of lifelong outdoor leisure activities.
They also counted the number and observed the degree of involvement of campers in this type of activity at camp. This helped them to document the people involvement.

The Laddin evaluators used informal interviews to discover the campers' reactions (opinions of, satisfaction with, and suggestions for) the actual outdoor leisure activities at camp.

KASA change represents changes in Knowledge, Attitudes, Skills, and Aspirations to behave in a certain way. (Note that aspirations to behave are different from actual behavioral change.) Camp Laddin evaluated this by using an questionnaire on attitudes about and intentions to participate in different outdoor leisure activities. They compared the responses of children who had attended to those of children who had not attended the camp.

The information about attitudes was interesting, but the camp's board of directors wanted to know if there were any demonstrable changes in the campers' behavior after leaving the camp. The evaluators investigated practice change by talking to parents on the telephone to determine if the campers had increased their participation in any of these activities after they returned home.

The final challenge for Camp Laddin was to assess the end results. This encompasses all of the lower levels and deals with broad changes in society. The camp evaluators set up a long-term plan to gather and analyze data (information) on the community's participation in and enjoyment of outdoor lifelong leisure activities.

To evaluate its success, Camp Laddin could have chosen to look at one or all of these levels. They could have built a thorough, efficient evaluation by combining several evaluation techniques in order to gather information on the levels of the evaluation which were most important to their evaluation audience. Notice that evidence of effects at the higher steps (i.e., KASA change, practice change, and end results) would be more convincing as proof of the program's impact. But, there is a catch: Assessment at the higher levels, although more convincing, is more costly and difficult. Based on the needs of your audience (those people who will use the evaluation results), you must make the decision about which level or levels of assessment will be most efficient for your program.
STATEMENTS OF PURPOSE

Writing a statement of purpose:

A statement of purpose should include the object(s) to be evaluated and the intended use of the information gathered. The following are examples of statements of purpose. The bracketed information provides an explanation of the parts of each statement.

• The purpose of this evaluation is to determine the effectiveness of the program in increasing participants’ self-control in order to help the Department of Health and Human Services better understand the program’s impacts. [The evaluation object is a specific program objective, that of increasing self-control. The intended use of the information is to determine the value for the Department of Health and Human Services, the evaluation audience.]

• The evaluation of Little Pony Camp will assess the instruction provided in the classes Equestrian Arts 1, 2, and 3. The information produced will be used by the equestrian instructors to improve their classes. [The evaluation object is the actual events of the program. The information is intended to be used by instructors, the evaluation audience, for program improvement.]

Notice that these examples included both an object and an intent which are specific for their program. The general format of the above purpose statements is easily adapted to many types of purposes.

Criteria:

With the objects of evaluation and the general purposes for evaluation considered, the next step is to write a statement of purpose for your evaluation. Before you begin, consider the following guidelines: Every evaluation purpose should be clear and understandable to all parties involved. The purpose should be relevant to the needs and goals of the program. The purpose should be feasible in terms of the time, resources, and expertise that you have available. The purpose should be useful to the evaluation audience, thus promoting use of the gathered information (adapted from Brinkerhoff, Brethower, Huchyj, & Nowakowski, 1983).

A statement of purpose is a very general statement. Through using the rest of the guidebook, you will "operationalize" your statement of purpose. That is, you will make it usable by specifying exactly how you can collect information on the object in order to fulfill the intent. With these thoughts in mind, begin the worksheets about purpose statements on the following pages.

A definitive purpose statement for the evaluation will provide guidance and direction.
On Clipboard #1 you listed the people who will be the primary audience(s) for your evaluation. Use this clipboard to select an evaluation object(s) and purpose(s), based on the needs of your evaluation audience.

First choose the intended use for the information provided by the evaluation by placing a check mark by one of the following choices.

- Program improvement
- Determining program worth of value
- Both/Other

Next, choose the object(s) of your evaluation from one of the categories previously described, or write in one of your own. Use check marks to indicate the object(s) of your evaluation. Then, select the areas of concern for each evaluation object.

<table>
<thead>
<tr>
<th>OBJECTS</th>
<th>AREA OF CONCERN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) PROGRAM</td>
<td>Actual events</td>
</tr>
<tr>
<td></td>
<td>Behaviors</td>
</tr>
<tr>
<td></td>
<td>of participants</td>
</tr>
<tr>
<td></td>
<td>of staff</td>
</tr>
<tr>
<td></td>
<td>of administrators</td>
</tr>
<tr>
<td></td>
<td>of others</td>
</tr>
<tr>
<td></td>
<td>Thoughts and attitudes</td>
</tr>
<tr>
<td></td>
<td>of participants</td>
</tr>
<tr>
<td></td>
<td>of staff</td>
</tr>
<tr>
<td></td>
<td>of administrators</td>
</tr>
<tr>
<td></td>
<td>of others</td>
</tr>
<tr>
<td></td>
<td>Other</td>
</tr>
</tbody>
</table>
### 2) PROGRAM PURPOSE
- External criteria
  - outside organization
  - parents
  - others
- Internal criteria
  - staff
  - program administration
  - board members
  - others

### 3) FACILITIES
- Satisfaction with the facilities of the staff, participants, parents, school boards, etc.
  - Cleanliness
  - Safety
  - Appropriateness for program activities
  - Environmental impact
  - Accessibility
  - Other

### 4) STAFF
- Performance
- Competencies
- Other

### 5) ADMINISTRATION
- Organizational structure
  - efficient
  - effective
  - other
- Policies and Procedures
  - pro environment
  - inclusive
  - meet legal requirements
  - other
6) SPECIFIC PROGRAM OBJECTIVE  Choose one of the three categories (affective, behavioral or cognitive), the specific topic of interest: and the group of people to whom it applies

<table>
<thead>
<tr>
<th>Affective Change</th>
<th>Camper</th>
<th>Staff</th>
<th>Parent</th>
<th>Teacher</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude toward the program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude toward the environment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude toward peers (social climate)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude toward religion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude toward school, learning...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Locus of Control (LOC)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-esteem</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Behavior change</th>
<th>Camper</th>
<th>Staff</th>
<th>Parent</th>
<th>Teacher</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cooperation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leadership</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem-solving</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduction of aggression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Maturity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cognitive change</th>
<th>Camper</th>
<th>Staff</th>
<th>Parent</th>
<th>Teacher</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific subject (i.e., biology, reading)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Now that you have identified the audience for your evaluation, chosen a use for the information, and selected the object(s) to evaluate, you have a basis for writing a specific statement of purpose for your evaluation.

Review the criteria for statements of purpose:
- clear and understandable
- relevant and compatible with program goals
- useful to the evaluation audience
- feasible

Write your statement of purpose:

_____________________________________________________________________________________________________
_____________________________________________________________________________________________________
_____________________________________________________________________________________________________

Does your statement of purpose satisfy the criteria listed above? Yes No

(Ask two other people who are familiar with your program to assess your statement of purpose using the criteria above.)

This statement of purpose will be the basis for your evaluation. As you complete Clipboards #3, 4, 5, and 6, refer back to this statement to guide your decisions.
To identify appropriate sources of data, consider:

1. your evaluation purpose,
2. relevant groups of stakeholders or organizational records, and
3. your access to the sources of data.

If you want to find out how good a new movie is, you might ask someone for his/her opinion. You probably will not ask just anyone, but someone who has seen the movie and who has similar tastes to yours. If you are on a limited budget, you will be even more cautious and will ask several people so that your can make the wisest use of your limited funds. When you ask for the opinions of several people, you are using different sources of data to evaluate the movie.

Most program evaluations also use multiple sources of data. Each stakeholder may have had a different experience with the program. If the evaluator can gather many viewpoints and "average" them together, then a realistic portrayal of the program can be created. In technical terms, the people or documents from whom you gather information are sources of data and may include any of the stakeholders or records of the program.

SOURCES

The purpose is the backbone of the evaluation. Fulfilling the purpose requires nutrients (information). The sources of data are the types of food that provide these nutrients. When you crave salt, fruit will not necessarily satisfy your needs. Similarly, the sources of data must match the evaluation needs of the particular program evaluation. You must choose appropriate sources of information for your evaluation based on the purpose of your evaluation. The sources of data will be unique for each evaluation.

Initially, you may want to consider all of your stakeholders and program records as sources of data. Your evaluation purpose can be your primary guide for selecting appropriate sources from all of the possible sources of data. It is often helpful to brainstorm all of the possible people and records that could be helpful in finding information related to the purpose. This will provide you with a list of potential sources. You can then refine the list based on your resources and your access to each source.

Certain sources of data may be particularly relevant for specific evaluation purposes. For example, in order to determine if the facilities are accessible to all, an evaluator might use personal opinions about the buildings and equipment. To gather opinions, the evaluator needs to pinpoint those people who use or may use the facilities. Then, the evaluator can select program participants, potential participants (including
those with disabilities who have a vested interest in accessibility), and staff as useful sources of data as opposed to school board members who may never have visited the site. Figure 3 suggests possible sources of data for each purpose listed in the previous chapter.

**FIGURE 3. Typical Sources of Data for Each Evaluation Purpose**

<table>
<thead>
<tr>
<th>EVALUATION PURPOSE</th>
<th>POSSIBLE SOURCES OF DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PROGRAM</strong></td>
<td></td>
</tr>
<tr>
<td>Actual Events</td>
<td>• Outside observers (consultants, professional peers)</td>
</tr>
<tr>
<td></td>
<td>• Staff observers</td>
</tr>
<tr>
<td></td>
<td>• Participants</td>
</tr>
<tr>
<td></td>
<td>• Visiting teachers</td>
</tr>
<tr>
<td></td>
<td>• Parents</td>
</tr>
<tr>
<td>Behaviors</td>
<td>• Outside observers</td>
</tr>
<tr>
<td></td>
<td>• Staff</td>
</tr>
<tr>
<td></td>
<td>• Participants</td>
</tr>
<tr>
<td>Thoughts and Attitudes</td>
<td>• Staff</td>
</tr>
<tr>
<td></td>
<td>• Participants</td>
</tr>
<tr>
<td></td>
<td>• Administrators</td>
</tr>
<tr>
<td><strong>PROGRAM PURPOSE</strong></td>
<td></td>
</tr>
<tr>
<td>External Criteria</td>
<td>• Professional experts</td>
</tr>
<tr>
<td></td>
<td>• Principals</td>
</tr>
<tr>
<td></td>
<td>• Visiting teachers</td>
</tr>
<tr>
<td></td>
<td>• Parents</td>
</tr>
<tr>
<td></td>
<td>• Participants</td>
</tr>
<tr>
<td>Internal Criteria</td>
<td>• Program administrators</td>
</tr>
<tr>
<td></td>
<td>• Board of Directors members</td>
</tr>
<tr>
<td><strong>FACILITIES</strong></td>
<td></td>
</tr>
<tr>
<td>Satisfaction, Cleanliness, Safety, Appropriateness for activities, Environmental impact, and Accessibility</td>
<td>• Parents</td>
</tr>
<tr>
<td></td>
<td>• Participants</td>
</tr>
<tr>
<td></td>
<td>• Visiting teachers</td>
</tr>
<tr>
<td></td>
<td>• Outside observers</td>
</tr>
<tr>
<td></td>
<td>• Staff</td>
</tr>
<tr>
<td></td>
<td>• Certifying organizations</td>
</tr>
</tbody>
</table>
FIGURE 3. Typical Sources of Data for Each Evaluation Purpose (Cont'd)

<table>
<thead>
<tr>
<th>EVALUATION PURPOSE</th>
<th>POSSIBLE SOURCES OF DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAFF</td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>• Participants</td>
</tr>
<tr>
<td></td>
<td>• Visiting teachers</td>
</tr>
<tr>
<td></td>
<td>• Outside observers</td>
</tr>
<tr>
<td></td>
<td>• Administrators</td>
</tr>
<tr>
<td></td>
<td>• Staff (self and peers)</td>
</tr>
<tr>
<td>Competencies</td>
<td>• Participants</td>
</tr>
<tr>
<td></td>
<td>• Visiting teachers</td>
</tr>
<tr>
<td></td>
<td>• Outside observers</td>
</tr>
<tr>
<td></td>
<td>• Administrators</td>
</tr>
<tr>
<td></td>
<td>• Staff (self and peers)</td>
</tr>
<tr>
<td>ADMINISTRATION</td>
<td></td>
</tr>
<tr>
<td>Organizational Structure</td>
<td>• Visiting Teachers</td>
</tr>
<tr>
<td></td>
<td>• Outside observers</td>
</tr>
<tr>
<td></td>
<td>• Administrators</td>
</tr>
<tr>
<td></td>
<td>• Board of Directors members</td>
</tr>
<tr>
<td></td>
<td>• Staff</td>
</tr>
<tr>
<td>Policies and Procedures</td>
<td>• Visiting Teachers</td>
</tr>
<tr>
<td></td>
<td>• Outside observers</td>
</tr>
<tr>
<td></td>
<td>• Administrators</td>
</tr>
<tr>
<td></td>
<td>• Board of Directors members</td>
</tr>
<tr>
<td></td>
<td>• Staff</td>
</tr>
<tr>
<td>SPECIFIC PROGRAM OBJECTIVES</td>
<td></td>
</tr>
<tr>
<td>Affective Change</td>
<td>• Participants</td>
</tr>
<tr>
<td></td>
<td>• Parents</td>
</tr>
<tr>
<td></td>
<td>• Visiting teachers</td>
</tr>
<tr>
<td></td>
<td>• Community resources (therapists, Scout leaders, neighbors, etc.)</td>
</tr>
<tr>
<td></td>
<td>• Staff</td>
</tr>
<tr>
<td></td>
<td>• Outside observers</td>
</tr>
<tr>
<td>Behavioral Change</td>
<td>• Participants</td>
</tr>
<tr>
<td></td>
<td>• Parents</td>
</tr>
<tr>
<td></td>
<td>• Visiting teachers</td>
</tr>
<tr>
<td></td>
<td>• Community resources (therapists, Scout leaders, neighbors, etc.)</td>
</tr>
<tr>
<td></td>
<td>• Staff</td>
</tr>
<tr>
<td></td>
<td>• Outside observers</td>
</tr>
<tr>
<td>Cognitive Change</td>
<td>• Participants</td>
</tr>
<tr>
<td></td>
<td>• Parents</td>
</tr>
<tr>
<td></td>
<td>• Teachers</td>
</tr>
</tbody>
</table>

The key to finding useful data is to ask the appropriate people or to search the most insightful records.
POSSIBLE SOURCES

Use Figure 3 to match your program purposes with possible sources of data. (Refer back to your statement of purpose on p. 30.)

In column 1, list all groups of people and types of records that may be useful in finding information for your evaluation purpose(s).

In column 2, list when you have access to each of the sources in column 1 (i.e., before, during, or after the program).

<table>
<thead>
<tr>
<th>Column 1</th>
<th>Column 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Possible Sources of Data</td>
<td>Access to Sources of Data</td>
</tr>
<tr>
<td>________________________</td>
<td>________________________</td>
</tr>
<tr>
<td>________________________</td>
<td>________________________</td>
</tr>
<tr>
<td>________________________</td>
<td>________________________</td>
</tr>
<tr>
<td>________________________</td>
<td>________________________</td>
</tr>
<tr>
<td>________________________</td>
<td>________________________</td>
</tr>
<tr>
<td>________________________</td>
<td>________________________</td>
</tr>
<tr>
<td>________________________</td>
<td>________________________</td>
</tr>
<tr>
<td>________________________</td>
<td>________________________</td>
</tr>
<tr>
<td>________________________</td>
<td>________________________</td>
</tr>
<tr>
<td>________________________</td>
<td>________________________</td>
</tr>
<tr>
<td>________________________</td>
<td>________________________</td>
</tr>
</tbody>
</table>

Through using Clipboards 1, 2, and 3, you have identified: 1) who will use the evaluation (your audience), 2) why you are conducting an evaluation (your purpose), and 3) where you can obtain the information (your sources of data).
DESIGN

In choosing a design for your evaluation, consider:

(1) your intent for the evaluation,
(2) your approach to planning, and
(3) the evaluation model that you prefer.

Thus far, you have identified the audience(s), purpose(s), and source(s) of data for your evaluation. You are now ready to design your evaluation. A design is an outline of the entire evaluation process. There are a number of issues that must be considered as you create your design; each will be addressed in this chapter.

1) CHOICES OF INTENT

If your intent is to evaluate for program improvement by the program staff, you will probably want to design an on-going evaluation that will provide continuous feedback. This will also allow you to see the short-term effects of specific program changes. This type of evaluation is called a formative evaluation and can be compared to a quiz designed to monitor your progress. If you are more interested in determining the value of the program (perhaps to determine if it is cost-effective to continue offering a white water rafting trip), then you are more likely to conduct a summative evaluation. Summative evaluations are similar to a final examination. They are an in-depth, over-all appraisal of the program and are most often conducted at the end of the event (Scriven, 1967). Choosing to conduct a formative, a summative, or both types of evaluation early in the planning process is important. Before making this decision, consider the needs of the evaluation audience and your purpose(s) for evaluating. Some characteristics of formative and summative evaluations are listed in Figure 4.
### FIGURE 4. Comparison of Formative and Summative Evaluations


<table>
<thead>
<tr>
<th>Formative</th>
<th>Summative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus is on program areas which need improvement, often identified by the staff.</td>
<td>Focus is on &quot;success&quot; or &quot;failure&quot; of certain aspects of or of the entire program.</td>
</tr>
<tr>
<td>It occurs in conjunction with the program, as a continual process.</td>
<td>It often occurs at the conclusion of the program, acting as a final test.</td>
</tr>
<tr>
<td>Typically emphasizes quick collection and interpretation of data for immediate use by the staff.</td>
<td>Information for making decisions about funding or support for the program is sought.</td>
</tr>
<tr>
<td>Program staff are usually the evaluation audience.</td>
<td>The evaluation audience is typically external to the program.</td>
</tr>
<tr>
<td>Multiple incremental improvements may be the results.</td>
<td>Justification for significant improvements may be produced.</td>
</tr>
<tr>
<td>External validity* may be sacrificed for usefulness.</td>
<td>Reliability and validity* are of paramount importance.</td>
</tr>
</tbody>
</table>

* For an explanation of validity and reliability, see Appendix 1.

### 2) CHOICES OF STYLE

When you plan your day, do you write out every task and devote to it a certain time, or do you just jump in and work on the most pressing demand until it is finished, and then decide what to do next? We probably all do some of each, depending on our personality and the nature of the work. Evaluators call this first style, the one with a detailed plan, a *fixed design*. With this type of design, you write a systematic plan at the beginning of the process and follow it without any major variation. The other style, which requires less in the way of initial planning, is called an *emergent design*. These designs evolve as the evaluation proceeds, so that decisions are made along the way based on what happens. Figure 5 further defines these styles. Think about the advantages and disadvantages of each for your situation.
Most evaluations fall somewhere between a "true" fixed and a "true" emergent design, incorporating a bit of both. The style for your evaluation will be a blend of your audience's needs, your personal preferences, and other aspects of the evaluation design (i.e., some types of design require more pre-planning than others). Understanding what your audience expects in terms of types of reports and amount of control over the evaluation planning process will help you adjust your style to their needs.
3) CHOICES OF MODELS

Many models of evaluation of educational programs have been developed. These models are like tent frames. There are many types of tent frames (i.e., pup or dome), yet they all do the same basic job (namely, giving structure and form to the tent material). If you are considering buying a tent, the type of frame that you select will depend on your budget and weight constraints, which tent feels more spacious to you, etc. Similarly, you should choose an evaluation model that "fits" you.

Evaluation models are usually theoretical conceptions of what an evaluation should include and how it should be conducted. It may be helpful to view these models as “persuasions” (Borich, 1983) rather than as a dictum of how to conduct an evaluation. All of the models were designed to meet certain types of evaluative needs. For example, some evaluation models focus on gathering information about the outcomes of the program (the reaction of the participants, the changes in behaviors, etc.). Other models focus on the processes that occur at the program. And, some models address the inputs of the program. Many combinations of these emphases are present in established models of evaluation. Each combination is well suited for specific evaluation needs.

The following are some of the commonly used models (Case Studies, CIPP Decision Making, Dialectic, Expert Opinion, Importance/Performance, Goal Attainment, Goal Free, Matrix, Naturalistic, Quasi-Experimental, and Responsive). You may want to use one, adapt one or combine several to create a framework with which you are comfortable. Only a brief overview is provided for each. Following the chapter, Figures 9 and 10 present a summary of the characteristics of each evaluation model. Consult the references provided for a fuller understanding.

Case Studies

If you want participants to learn about wildflowers, would you have them study every flower, or even most of the flowers, in a field? Probably not. Most of us would agree that by focusing on one or a few, a person could get a pretty good idea about the form, function, and beauty of flowers. This type of exercise is actually a case study.

An evaluator using a case study model is like a detective who tries to discover and understand the interactions, experiences, and effects of an entire program on one or several representative participants. Instead of studying the effects on all of the participants, (in which case the evaluation could not be in-depth), case studies examine the intricate and subtle effects of the program on a limited number of participants. From this type of evaluation, you can learn about the unique experiences of individuals. Case studies may have a fixed design, but more often they utilize an emergent design. These evaluations rely on many different techniques and sources of data for gathering information. This variety helps to insure that the information which is collected is accurate. Figure 6 provides an example of a plan that might be developed for a case study.
FIGURE 6. Case Study Plan

<table>
<thead>
<tr>
<th>A. Before the visit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) preliminary personal interviews with participant, parents, and behavioral therapists to gather expectations and to set personal goals;</td>
</tr>
<tr>
<td>2) pretests;</td>
</tr>
<tr>
<td>3) observations in the home setting.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. During the visit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) on-site observation and rating of behaviors;</td>
</tr>
<tr>
<td>2) interviews to discuss experiences and progress toward goals;</td>
</tr>
<tr>
<td>3) collection of journal writings to be analyzed for attitude changes.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C. After the visit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) telephone interviews with parents and therapists to determine changes in behavior;</td>
</tr>
<tr>
<td>2) posttests;</td>
</tr>
<tr>
<td>3) follow-up questionnaires sent several months after the visit.</td>
</tr>
</tbody>
</table>

Imagine the large time commitment required for the example above. This design model is intense, but the rewards are well-supported information about the processes, and the expected and unexpected outcomes of the program. Another advantage is that case studies give you a portrayal of the program from the participant's first contact with the program to its lasting effects.

Case studies have been criticized on the basis of the tiny sample studied. If a representative sample is not chosen, the information that is gathered may be of limited value. However, this type of personal information is likely to be important and persuasive to stakeholders who have a strong interest in the personal growth of individuals (i.e., parents, and staff). Case studies are especially appropriate for program monitoring or description and for determining the level of achievement toward certain objectives. Such evaluations typically provide in-depth data about what actually goes on (the processes) and the effects of the program (the outputs).

References:

General:


Examples:


CIPP Decision-Making

CIPP stands for Context (goals and objectives, needs, planning decisions), Input (plans, resources, structuring decisions), Process (activities, procedures, implementing decisions), and Product (outcomes, continuation decisions). If you are looking for a comprehensive, structured approach, the CIPP model might be a good choice. This model is popular because it is logical and based on decision-making needs. It will guide you through a comparison of the actual activities and results with those that are desired. A CIPP evaluation traces a program from its beginning to end, incorporates both formative and summative evaluations, frequently uses a fixed design, and yields information that can be used for making decisions about the efficiency of the program. This design may be ideal for evaluating a program's organizational design.

The complexity and time-consuming nature of the CIPP model may be a drawback of this model. It may not be efficient to conduct a formal evaluation at all of the stages of program development. Additionally, this model assumes that the informational needs of decision makers can be clearly identified in advance; yet, this is not always possible.

References:

General:


Examples:


Dialectic

Have you ever listened to two people argue over an issue and really learned something from their debate? Debate can be an effective learning tool because it stretches our minds and forces us to search out information in order to prove our point. Some evaluators have taken advantage of these characteristics of logical argumentation, or dialectics, by incorporating a form of debate into their evaluation.

For example, a transactional evaluation uses input from two parties, from those who would be helped by a change in the program and from those who would be hurt. This type of evaluation is commonly used to assess the benefits and costs of a new program.

An adversary evaluation uses two evaluation "teams." One team is commissioned to support the program, demonstrating its positive impacts, and the other team is to prove the negative aspects of the program. The teams present their cases to a decision maker who then makes a final ruling.
Another approach that uses a form of debate to build consensus is the Delphi technique. A panel of interested stakeholders provides anonymous responses to several rounds of questionnaires. These require identification and ranking of the program goals, needs, etc. On subsequent rounds, the panelists are provided the averaged results from the previous questionnaire round. If someone disagrees with the majority, then that person can re-do his or her ranking, or can try to convince the others that his or her initial ranking is better than theirs. This process can continue for several rounds until there is a relatively uniform agreement. The result is a rank ordered list of concerns or priorities as perceived by stakeholders.

Dialectic models can uncover a large amount of information and may safeguard against one-sided evaluations. The primary drawback of dialectic evaluations is that they require an evaluator who is skilled in interpersonal relations. They also require a number of panelists or evaluation team members who are willing to make a commitment of time and energy, and dialectic models do not provide immediate feedback. These models do, however, provoke in-depth consideration and may be most appropriate for clarifying and prioritizing a program's purposes. Dialectic models may also be especially helpful to evaluators who want to determine the overall worth of a program.

References:

General:

Transaction:


Adversary:


Delphi:


Expert Opinion

An expert is one who is highly trained in a specific field of knowledge. It is reasonable to assume that an expert's judgment of a program could be useful in determining the value of a program and that suggestions which an expert makes could be helpful for improvement. In the field of camping, experts are often used as consultants to evaluate a program in terms of health care, physical facilities, and safety procedures. "Experts" take the form of independent consultants, professional peer reviews, standards committees (such as that established by the American Camping Association (ACA) for accreditation), or program initiators. One or all three types of information (input, process, or output) may be provided from these evaluations that are commonly used for evaluations of staff and facilities.

References:

General:

Examples:


Goal Attainment

Traditional education evaluation models begin with established goals or objectives and seek to determine the program's progress toward these goals. Goal attainment models utilize a summative approach and usually include testing of individuals to determine the amount of behavior, knowledge or attitude change that has been achieved. One risk of this approach is that it may ignore the process used to reach the goal and may fail to address unintended effects. For example, funders may be very satisfied with results that show that a group of teenagers really enjoyed their time at camp. Would they, however, be so happy if they knew that the group's fun came from dominating younger campers and that the effects on some of the younger campers were traumatic? If you rely solely upon this model, then you assume that achievement of your primary goal or objective is more important than determining any of the other effects on the participants.

One advantage of this type of evaluation is that the evaluation questions and purpose are straightforward and easy to understand. Thus, the majority of the evaluation resources can be spent on collecting and analyzing the data. Primarily, goal attainment evaluations provide information about program output, and they are particularly appropriate for evaluations of specific program objectives or of the personal goals of staff and participants.

References:

General:


Example:

Goal-Free

If you take a walk in the woods, knowing that it just rained, you would expect to see lots of mushrooms. Do you think that you would notice more of them than if you were not expecting to see them? Sometimes our expectations color what we see. Could this suggest that in conducting an evaluation, with a defined purpose or goal, we see, hear, and value only expected outcomes? Some evaluators feel that traditional, goal-based evaluations are innately biased, ignoring both positive and negative unexpected effects. With a goal-free design, the evaluator strives to shrug off preconceptions and to take a fresh, untainted look at the program. This is ideal for discovering what is actually happening in the program and can yield input, process, and/or output data. However, this model has been faulted for being impractical. Since there is no focus for the data collection, the evaluator could spend much time collecting irrelevant information. Also, it is impossible for an evaluator to cast off of all preconceptions about a program.

Although using the goal-free evaluation model may not be feasible for a formal evaluation, it may be appropriate to incorporate the ideals of this model into your evaluation. This approach is often used as a preliminary evaluation step in order to determine 'what's out there.' For example, an evaluator who is interested in the program's effects on the participants' behaviors may first visit the program noting all of the behaviors that can be observed. The evaluator may then develop a behavior rating scale which limits observation to specific behaviors which are of interest to the evaluation audience.

References:
- Walsh, P.L. (1980). An empirical evaluative comparison of the goal-based and goal-free approaches to educational evaluation (Published doctoral dissertation, University of Illinois at Urbana).
Importance/Performance

This model will be especially useful if you are interested in evaluation for marketing purposes or in determining the relative importance of various program areas. Importance/performance evaluations gather information about how important certain elements of the program are and the quality of the level of performance (i.e., How important are the swimming lessons and how well are the children learning to swim?). You can plot the results on an x-y axis and then easily identify areas of high importance and low performance, indicating the need for improvement in that area. (See Figure 7 below.) The importance/performance model focuses on evaluation of output information and is particularly useful for assessing the achievement of and for prioritization of program objectives.

![Figure 7: Importance/Performance Results](image)

**References:**

Camping:


**Matrix**

Some models use matrices to build a framework for their evaluation. To better understand a program, we can think of it as several parts joined together in a series. A program may consist of: (1) antecedents (staff training, materials, initial attitudes of participants), (2)
transactions (teaching methods, activities, interactions among students), and (3) outcomes (knowledge learned, damage to equipment, lasting impressions). An example of part of a matrix is given in Figure 8 below. If you were using this matrix to evaluate a canoeing activity, you could first describe the expected inputs, processes, and outcomes. You could compare how these fit together; is it logical to expect participants to learn the skills from the planned activities? Then you could observe and describe the actual inputs, processes, and outcomes. You could then consider whether or not the actual connections were logical. Finally, through comparing the expected to the actual inputs, processes, and outcomes, you could discover where the program is not meeting the expectations. Matrices may be excellent for guiding your evaluation. However, be aware that a matrix may act as a blinder if it is too narrowly focused.

FIGURE 8. Example of a Matrix for the Evaluation of a Canoeing Activity

References:
Naturalistic

Some people believe that to truly understand outdoor education programs, we must look at the program as a part of an ever-changing reality. It is as if the "reality" of the situation is a soap bubble. A bubble is always moving and changing. It looks different from every angle, and if you look at it twice, it will never be the same. Outdoor education programs and the social interactions that occur in these programs are ever-changing, look different to different people with different perspectives, and no two programs are ever the same. Evaluators who take this "naturalistic" point of view, have developed techniques and models to guide their evaluations. Naturalistic evaluations are like impressionistic paintings. They use colorful bits of information from many perspectives to create a collection of descriptions (dots of color, if you will) which give the audience a vibrant view of the entire program. These models emphasize studying the program as an interconnected whole rather than as independent parts. Naturalistic evaluations use emergent designs and use in-depth descriptions, as opposed to statistics, to report their findings. Naturalistic evaluators try to see the program through the participants' eyes and use many methods to collect data from different viewpoints.

Here is an example of how a naturalistic evaluator might assess the effects of a program on the participants' self-esteem. The evaluator might become a participant in the activities in order to obtain an insider's perspective. Information might be collected through personal perceptions, observations, and informal discussions with the participants. After initial activities and interactions with the participants, the evaluator would summarize impressions into a working hypothesis (or statement of explanation) which the evaluator would continue to modify and support as more information was gathered. The major concern would be in describing the situation fully to gain a more complete understanding of the many factors which affect the participants' self-esteem. Close contact with the group leaders would provide continual feedback. Eventually, the evaluator would write a detailed report with the intent of including all information that could be useful to the evaluation audiences.

One of the major disadvantages of naturalistic models is that some audiences do not trust verbal or written descriptions. If your audience expects statistics to support your findings or wants proof of cause and effect, then this would not be a good choice of design for you. However, this model is an excellent choice for those interested in the processes occurring during the program and the program's intended and unintended effects. It is well suited for the purposes of program monitoring and description.

References:

General:


Examples:

Quasi-Experimental

If you are interested in determining cause and effect or in studying the effects of the entire program, a game, a trip, or a teaching method, then you may want to use a quasi-experimental approach. A "true" experiment would require a random selection of participants, rigorous control over the "treatment" (the program), and a comparison of people who had and had not been exposed to the program. Since true experiments are not usually feasible, often quasi-experiments are conducted. Quasi-experiments maintain some of the rigor and the comparison characteristics, but lack random assignment of individuals to randomly selected programs. The comparison can be made between a group of people who were exposed to the program, game, etc. (i.e., the treatment group) and a group who were not exposed to the treatment (i.e., the control group). Alternatively, a comparison can be made of the treatment group before the experience to the same group after the experience. Careful measurement and control of the situation should be emphasized. Reliability and internal and external validity are essential. (These are further discussed in Appendix 1.) Typically, quasi-experimental designs utilize tests and written questionnaires for data collection. The time at which data are collected is particularly important for these types of evaluations. Refer to Appendix 2 for an explanation of the most common data collection times. Quasi-experimental models typically are summative in nature and have a fixed design.

The following is an example of a quasi-experimental evaluation designed to determine the effects of a program on participants' self-esteem. The evaluator might begin with a hypothesis (a statement about what is expected to be found) that states: participants' self-esteem will increase as a result of a white water rafting adventure vacation. The evaluator would form this hypothesis from reading the results of past studies. The next step would be to write a comprehensive plan for the evaluation. This would be constructed so that no other factors could affect the participants' self-esteem. (Examples of these other factors, called intervening variables, are: other challenge activities or the participants knowledge that they are being studied.) The evaluator may measure self-esteem through a test and report results in "hard" data (numbers, percentages) following established statistical procedures. Much time would be spent creating and administering the test (instrument) to assure accurate and consistent measures of self-esteem. A comparison of the participants' responses prior to and after the activities could demonstrate that the activities caused the change in self-esteem.
The major disadvantage of quasi-experimental designs is the artificial conditions under which people may not act naturally. Imagine this: a team of evaluators is examining the effects of a ropes course experience on the group cooperation among participants. To do this, they compare five groups who experience the course to five groups who do not. As the groups complete the course, all must have the same instructions, use the same elements, and be given the same debriefing. This type of control is contradictory to the purposes of a ropes course. Most groups are different and require specialized instructions, levels of challenges, and post-experience discussions. Do you think the experimental control would change the effects of the course? Many believe that it could have drastic effects. If you choose this type of design, be cautious about these effects.

The major advantage of quasi-experimental designs is that they provide objective data that is perceived to be reliable and can be replicated in future studies. Quasi-experimental designs are well suited for documenting attitude, behavioral, or cognitive change as they typically provide information on the program's outputs.

References:

General:


Examples:

• Becker, L.M. (1977). The effect of resident outdoor experience on attitudinal change toward environmental issues. (Doctoral dissertation, University of Northern Colorado). (University Microfilms No. 77-30, 797)


Responsive

A widely accepted model for evaluation is the responsive model. This model is oriented toward the needs of the stakeholders. Evaluators which apply this model focus on the actual program as opposed to the intended program, and they are sensitive to the varying perspectives of different stakeholders. The evaluation is structured around the expressed interests of the stakeholders. This model requires an evaluator who can determine the stakeholders’ needs through observation and informal conversations and who can adapt to them. An effective evaluator is like an experienced counselor leading a sing-along. Every counselor may know many songs, but a practiced counselor can judge the audience and tell if an upbeat, silly song would work well, or if the audience needs a slow, melodic song. Responsive evaluators will pay attention to the needs of their audience and provide input, process, or output information as needed.

References:

Examples:

When choosing a model, consider the evaluation audience, your preferences in planning style, and the intent for the evaluation.

Use the references listed to further investigate your chosen model.

Figures 9 and 10 summarize the characteristics of these models.
**FIGURE 9. Generalized Characteristics of the Evaluation Models**

NOTE: These are generalizations and represent only the typical focus of each model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Intent (Formative, Summative, or Either)</th>
<th>Style (Emergent, Fixed, or Either)</th>
<th>Type of Information Provided (Input, Process, Output, or All)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASE STUDY</td>
<td>Either</td>
<td>Either</td>
<td>Process and Output</td>
</tr>
<tr>
<td>CIPP</td>
<td>Summative</td>
<td>Fixed</td>
<td>All</td>
</tr>
<tr>
<td>DIALECTIC</td>
<td>Either</td>
<td>Either</td>
<td>All</td>
</tr>
<tr>
<td>EXPERT OPINION</td>
<td>Either</td>
<td>Either</td>
<td>All</td>
</tr>
<tr>
<td>GOAL-FREE</td>
<td>Either</td>
<td>Emergent</td>
<td>All</td>
</tr>
<tr>
<td>GOAL ATTAINMENT</td>
<td>Summative</td>
<td>Fixed</td>
<td>Output</td>
</tr>
<tr>
<td>IMPORTANCE/PERFORMANCE</td>
<td>Either</td>
<td>Fixed</td>
<td>Process and Output</td>
</tr>
<tr>
<td>MATRIX</td>
<td>Either</td>
<td>Fixed</td>
<td>All</td>
</tr>
<tr>
<td>NATURALISTIC</td>
<td>Formative</td>
<td>Emergent</td>
<td>Process</td>
</tr>
<tr>
<td>QUASI-EXPERIMENTAL</td>
<td>Summative</td>
<td>Fixed</td>
<td>Output</td>
</tr>
<tr>
<td>RESPONSIVE</td>
<td>Formative</td>
<td>Emergent</td>
<td>All</td>
</tr>
</tbody>
</table>
### FIGURE 10. General Advantages and Disadvantages of the Evaluation Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASE STUDY</td>
<td>•Gathers intensive information from each stage of the program</td>
<td>•Time-consuming&lt;br&gt;•Small sample may not provide representative information</td>
</tr>
<tr>
<td></td>
<td>•Considers the experiences of individuals</td>
<td></td>
</tr>
<tr>
<td>CIPP</td>
<td>•Comprehensive&lt;br&gt;•Provides information for decision makers</td>
<td>•Time-consuming&lt;br&gt;•Complex</td>
</tr>
<tr>
<td>DIALECTIC</td>
<td>•Yields information from many viewpoints</td>
<td>•Requires a large evaluation team.</td>
</tr>
<tr>
<td>EXPERT OPINION</td>
<td>•Offers different perspectives and insights</td>
<td>•Relies only on the judgment of one or a few people</td>
</tr>
<tr>
<td>GOAL-FREE</td>
<td>•Addresses all of the program’s effects</td>
<td>•Evaluator must be trained and understand personal biases</td>
</tr>
<tr>
<td>GOAL ATTAINMENT</td>
<td>•Focuses on specific questions</td>
<td>•Ignores processes and unintended effects</td>
</tr>
<tr>
<td>IMPORTANCE/</td>
<td>•Prioritizes needs</td>
<td>•Prioritization may be difficult since most program elements are seen as important</td>
</tr>
<tr>
<td>PERFORMANCE/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MATRIX</td>
<td>•Comprehensive&lt;br&gt;•Structured</td>
<td>•Complex and demanding</td>
</tr>
<tr>
<td>NATURALISTIC</td>
<td>•Considers multiple influences&lt;br&gt;Studies the entire environment</td>
<td>•Time-consuming&lt;br&gt;•Unfocused&lt;br&gt;•Results may not be trusted</td>
</tr>
<tr>
<td>QUASI-EXPERIMENTAL</td>
<td>•Yields &quot;hard,&quot; objective data&lt;br&gt;Offers well-researched techniques are available&lt;br&gt;Provides intense study of one aspect of program</td>
<td>•Results may be artificial&lt;br&gt;•Requires a controlled atmosphere&lt;br&gt;•May ignore the effects of multiple influences</td>
</tr>
<tr>
<td>RESPONSIVE</td>
<td>•Provides for the evaluation audiences’ needs</td>
<td>•Requires knowledge of numerous evaluation techniques</td>
</tr>
</tbody>
</table>
YOUR DESIGN

Referring back to Clipboard #2 on p. 27, restate the intent for your evaluation.

___ Program improvement
___ Establishing worth or value
___ Both/Other

Considering the needs of your evaluation audience and your evaluation purpose, which type of design do you plan to use?

___ Formative
___ Summative
___ Both

Will access to your sources of data limit your choice of design? If so, how?

______________________________

Consider your approach to planning. Which planning style are you comfortable with?

___ Fixed design
___ Emergent design
___ Both

Using Figure 9 on page 50 and your answers to the previous questions, identify the evaluation models which are appropriate for your evaluation.

______________________________

After reading the brief descriptions of the evaluation models, do you feel that any of them fit your needs? Which ones? If none do, can you combine several or adapt one to create something more appropriate?

______________________________

52

Based on decisions that you identified on Clipboards #1, 2, and 3, you have chosen an evaluation design. Use the following sections and clipboards to adapt that design for use with your specific program.
In determining the available resources, consider:

1. time,
2. expertise (knowledge), and
3. monetary resources.

The resources are the muscle power behind the evaluation. Everyone expects an evaluation to require an investment of resources; however, the quality of the evaluation is not solely dependent upon these expenditures. Quality relies more on the dedication of the evaluator to find information that is useful and relevant to the needs of the audiences. Careful planning will make your evaluation successful as well as efficient.

The first step is to take an inventory of the available resources. When considering time, determine the number of staff hours as well as the time span that you and your staff can dedicate to the evaluation. (Time frame is further addressed in the next chapter.) In the category of expertise, search for staff, personal contacts, volunteers, university researchers, and fellow professionals who have knowledge or experience in evaluation. Evaluators use many skills. Some skills to look for in potential members of your evaluation team are: communications, interpersonal, organizational, computer, research, and statistical skills.

It is well worth it to plan evaluation expenses into your annual budget. This is much easier than trying to extract evaluation funds from other program areas; the feedback that you receive should be worth the expense. Be sure to consider how allotment of funds and staff time will affect your program and how this allotment will be accepted by staff and funders. For additional aid, consider asking for volunteers; many people who truly believe in the value of residential outdoor education are willing to help support their belief. An excellent source of resources may be the nearest university. Often, faculty and graduate students are looking for interesting research projects. They can provide information about current evaluation practices, and they possess many of the key skills listed above. A joint effort with a university can be mutually beneficial. You acquire access to expertise and personnel and the university gains a hands-on learning experience for students. Some level of funding may be required. Also, you may want to hire an intern to supply extra help if the time demands of the evaluation overburden your regular staff.

More Information:

"Interdisciplinary Opportunities in a Camp Setting: Work/Service/Research/ Education."
Mary Faeth Chenery, Department of Outdoor Education, LTUB, P.O. Box 199, Bendigo, Victoria 3550, Australia. Fax (61-54)447-777;

Careful planning is the key to conducting a successful evaluation on limited resources.
YOUR RESOURCES

Consider the amount of each resource that you have access to. Place an X on the line next to each category of resource to indicate the amount available for your evaluation.

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAFF HOURS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EVALUATION KNOWLEDGE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MONETARY RESOURCES</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Create a list of your evaluation team (those people who will design, conduct, or interpret the evaluation). List the people who may be available, the hours that each might devote to the evaluation, and the expertise that they can contribute.

<table>
<thead>
<tr>
<th>Team member</th>
<th>Hours</th>
<th>Area of Expertise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What effect will low or high resources (staff hours, knowledge, or monetary resources) have on your evaluation design? __________________________
In choosing your time frame, you should consider four factors:

1. your evaluation purpose,
2. the evaluation model that you have chosen,
3. the availability of the evaluation team, and
4. when and how you can get information from the sources of data.

When should you conduct your evaluation? How will you fit your data collection into your program? And, how long will the evaluation last? These are questions which pertain to the time frame of the evaluation. Although you are actually informally evaluating all of the time, your formal evaluation will probably be restricted to certain periods of time.

A written time frame can be a helpful planning tool. It should describe when each of the steps of the evaluation (development of evaluation design, data collection, analysis of data, report of results, application of results, and meta-evaluation) will be conducted in relation to the entire program and specifically when data will be collected relative to a session of the program. Common collection times are: before the experience (called a pretest), during the experience, at the end of the experience (a posttest), and an extended time after the experience (a follow-up). [Note: although these are termed, pre/post "tests," in this context they refer to any type of data collection, e.g., observation, surveys, behavior rating.]

1) EVALUATION PURPOSE

The first factor relating to time frame is your evaluation purpose. The timing of the evaluation process in relation to that of the program will depend upon your broad purpose(s) (for program improvement or for demonstrating the value of a program) and upon the specific evaluation purpose that you have written. An evaluation should be structured to provide information at a time when that information can be used. In other words, if your evaluation is designed to demonstrate the value of your program to a school board, then you need to provide that information before the school board makes a decision about sending students to your program. For program improvement purposes, you will probably want evaluation results to be available in time to be incorporated into the planning for the next session or season. In order to provide useful information, an evaluation must also provide timely information.
The specific purposes for your evaluation may necessitate data collection at a certain time or times. For example, if your purpose is to demonstrate a change in student behavior, you will want to collect data both before and after but not during the session. Conversely, for information on the actual program processes, the most accurate information can be gathered during the program. Some purposes, such as assessing facilities, allow you flexibility in that they can be conducted at any time.

2) EVALUATION MODEL

The second factor that may influence your choice of time frame is your evaluation model. Some of the evaluation models require data collection at all of the common collection times; some require collection at one specific time; and, others only recommend data collection times. For instance, to document a decrease in racial prejudices among a mixed group of adventure campers, an evaluator might choose a quasi-experimental design. The quasi-experimental model requires either pre/posttesting of an experimental or the use of a control group. If the evaluator failed to collect data at these times, adequate comparisons could not be made, and the quasi-experimental model could not be applied. Figure 11 lists various models and their required or recommended collection periods. Refer to this table to review the requirements of the model that you have selected.

FIGURE 11. Models and When Data are Collected

<table>
<thead>
<tr>
<th>MODEL</th>
<th>Pretest</th>
<th>During</th>
<th>Posttest</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASE STUDY</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>CIPP</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>DIALECTIC</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>EXPERT OPINION</td>
<td>--</td>
<td>S</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>GOAL ATTAINMENT</td>
<td>R</td>
<td>--</td>
<td>R</td>
<td>--</td>
</tr>
<tr>
<td>GOAL FREE</td>
<td>--</td>
<td>S</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>IMPORTANCE/PERFORMANCE</td>
<td>--</td>
<td>--</td>
<td>R</td>
<td>--</td>
</tr>
<tr>
<td>MATRIX</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>NATURALISTIC</td>
<td>--</td>
<td>R</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>QUASI-EXPERIMENTAL</td>
<td>R (or a control group)</td>
<td>--</td>
<td>R</td>
<td>--</td>
</tr>
<tr>
<td>RESPONSIVE</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
</tbody>
</table>

*R = Model requires collection at this time.
*S = Model suggests collection at this time.
3) EVALUATION TEAM

You have considered how much actual time your evaluation team can devote to the evaluation, but have you thought about when that time will be required? Do you have personnel available throughout the session (a requirement for some techniques such as direct observation of activities)? Can you and your staff set aside a large portion of time for planning and development of instruments? Will your summer program staff be available to collect follow-up data in six months? Without proper consideration of the amount of time, the time span, and how crucial parts of the evaluation fit into your staff's schedule, problems may arise. For example, you do not want counselors to be making observations and recording data when they are also supervising campers. This may result in poor data collection (or worse, poor supervision). Thus, the third important consideration for a well chosen time frame is the availability of the evaluation team members.

4) SOURCES OF INFORMATION

The fourth factor in the choice of a time frame is the sources of the information. What might happen if you plan to conduct pretests on participants, but you cannot contact participants early enough? Or, what if you expect campers to take time out from an exciting game to complete a long questionnaire? Both of these might fail because the evaluator ignored the time constraints on the sources of data. This may limit data collection, but if you are creative, you can find ways of accessing the stakeholders who can provide information. Some possible ways to contact hard-to-reach stakeholders include: telephone interviews of board members, video tapes of participants' behaviors when they are at home, and discussions with parents during Visitors Days. Be sure to investigate all of the possibilities.

In choosing your time frame, you should consider four factors:

1) your evaluation purpose,
2) the evaluation model that you are working with,
3) the availability of the evaluation team members, and
4) when you can gather information from the stakeholders.
PURPOSE

Based on your broad evaluation purpose(s), do you have deadlines by which you must have evaluation results? When do you want to have evaluation results?

__________________________________________

__________________________________________

Referring to your purpose statement (p. 30), which data collection time(s) might be appropriate for your evaluation?

____ Before the experience
____ During the experience
____ After the experience
____ Other ________________________________

MODEL

Based on your model of choice, which of the collection times would be appropriate? (Refer to Figure 11 on page 56.)

____ Before the experience
____ During the experience
____ After the experience
____ Other ________________________________

EVALUATION TEAM

When are the evaluation team members not available? What effects will this have on your choice of time frame?

__________________________________________

__________________________________________
ACCESS TO SOURCES OF DATA

On page 34 you listed your possible sources of data and times at which you have access to them. How might these limit your evaluation time frame?

PLANNING YOUR TIME FRAME

Combining all of your responses above, create a time line for your evaluation. Describe when the evaluation will be conducted in relation to the entire program (i.e., May - develop the design; June 19-July 3, July 10-July 23, and July 31-August 20 - collect data; September - analyze data; November 12 - report results; December-March - use results in the planning of the following year's program and evaluation). Also, describe specifically when data will be collected (i.e., before the experience while participants are in their home, the first night of the program, on the third day of the program during recreation, in the classroom after the experience, etc.).

<table>
<thead>
<tr>
<th>Date</th>
<th>Evaluation Event</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Section 2
Evaluation Techniques
TECHNIQUES

The first section of this guidebook has identified evaluation audiences, purposes, resources, sources of data, and time frames. The remainder of the guidebook is devoted to exploring a set of “techniques” or methods for gathering information. The techniques presented are those more frequently utilized in the evaluation research; this collection of techniques is not all inclusive. Often when we think of certain evaluation techniques, we associate them with a specific evaluation model (e.g., we often pair the use of tests to a quasi-experimental model). Yet, techniques do differ from models. The techniques are the tools that are used to gather information; the design is the framework or the outline used to guide the entire evaluation process. Often it is appropriate to build an evaluation with one design and several different techniques. Using multiple techniques is similar to looking at a rock from different angles. Each angle provides a different perspective and a better understanding of the rock. In the same manner, an evaluator may use a combination of participant observation, personal interviews, and artifact analysis to gain a multifaceted picture of a program. This use of multiple techniques to gather information about one evaluation object is called triangulation.

There is a tendency for evaluators to learn one evaluation technique and to apply it in every evaluation. This is called “the Rule of the Hammer” and was elaborated on in an article by White (1975). What happens when you give a young child a hammer? Usually, the child uses the hammer to pound everything. You can imagine the results. The overuse of an evaluation technique can produce similar results. Use this section to avoid “the Rule of the Hammer.”

This section contains descriptions of various data collection techniques. You can use this section to learn more about specific techniques that you have already identified as appropriate for your program evaluation. Section two can also serve as a reference for learning more about other techniques. For each technique, a brief description, advantages, disadvantages, and additional references are provided. The following major categories of techniques are included in this section:

1. behavior observations,
2. written questionnaires,
3. verbal techniques, and
4. miscellaneous techniques.

The following five pages and Figure 12 provide recommended techniques for each of the specific purposes which were previously discussed. You can use these charts along with your evaluation purpose (p. 30), the object of your evaluation (p. 27-29), and your sources of data (p. 34) to identify possible techniques.
If the **PROGRAM** is your evaluation object,

**ACTUAL EVENTS** are what interests you.

An **OUTSIDE OBSERVER** is your source of data.

Then try: Unobtrusive Observation, Direct Observation, Participant Observation, Check Sheet, Journal, or Anecdotal Recording.

A **STAFF** member is your source of data.

Then try: Unobtrusive Observation, Direct Observation, Participant Observation, Check Sheet, Journal, Anecdotal Recording, or Experience Sampling.

A **VISITING TEACHER** is your source of data.

Then try: Check Sheet, Journal, or Anecdotal Recording.

A **PARENT** is your source of data.

Then try: Anecdotal recording.

A **PARTICIPANT** is your source of data.

Then try: Check Sheet, Journal, Anecdotal Recording, or Experience sampling.
If the **Program** is your evaluation object,

**Behaviors** are what interests you,

An **outside observer** is your source of data,

Then try: Direct Observation, or Unobtrusive Observation.

A **staff** member is your source of data,

Then try: Direct Observation, Participant Observation, Unobtrusive Observation, Anecdotal Recording, or Self-Evaluation.

A **participant** is your source of data,

Then try: Direct Observation, Participant Observation, Unobtrusive Observation, Anecdotal Recording, or Self-Evaluation.

**Thoughts and attitudes** are what interest you,

A **staff** member is your source of data,

Then try: Artifact Analysis, Self-Evaluation, Written Questionnaire, Participant Observation, Anecdotal Recording, Personal Interview, or Experience Sampling.

A **participant** is your source of data,

Then try: Artifact Analysis, Self-Evaluation, Anecdotal Recording, Written Questionnaire, Participant Observation, Personal Interview, or Experience Sampling.

A **parent** is your source of data.

Then try: Anecdotal Recording, Written Questionnaire, or Personal Interview.

An **administrator** is your source of data.

Then try: Self-Evaluation, Written Questionnaire, or Personal Interview.
If the **PROGRAM PURPOSE** is your evaluation object,

**EXTERNAL CRITERIA** are to be used.

An **OUTSIDE EXPERT** is your source of data.

Then try: **Outside Observer.**

A **PRINCIPAL, VISITING TEACHER, PARENT, or PARTICIPANT** is your source of data.

Then try: **Q-Sort, Written Questionnaire, Focus Group, or Personal Interview.**

**INTERNAL CRITERIA** are to be used.

Then try: **Q-Sort, Written Questionnaire, Focus Group, or Personal Interview.**

If the **FACILITY** is your evaluation object,

Then try: **Written Questionnaire, Outside Observer, Anecdotal Recording, Personal Interview, Focus Group, or Check Sheet.**

If the **STAFF** is your evaluation object,

Then try: **Written Questionnaire, Personal Interview, Anecdotal Recording, Direct Observation, Self-Evaluation, or Test.**

If the **ADMINISTRATION** is your evaluation object.

Then try: **Focus Group, Written Questionnaire, Personal Interview, Outside Observer, Anecdotal Recording, or Check Sheet.**
If a **PROGRAM OBJECTIVE** is your evaluation object,

**AFFECTIVE CHANGE** interests you.

A **PARTICIPANT** or **VISITING TEACHER** is your source of data.

Then try: Experience Sampling, Sociometric Method, Self-Evaluation, Artifact Analysis, Games Observation, Participant Observation, Unobtrusive Observation, Standardized Indicator, Personal Interview, Written Questionnaire, Focus Group, or Anecdotal Recording.

A **PARENT** or **COMMUNITY RESOURCE MEMBER** is your source of data.

Then try: Standardized Indicator, Personal Interview, Written Questionnaire, Focus Group, or Anecdotal Recording.

A **STAFF** member is your source of data.

Then try: Experience Sampling, Sociometric Method, Self-Evaluation, Artifact Analysis, Standardized Indicator, Personal Interview, Written Questionnaire, Focus Group, Anecdotal Recording, Participant Observation, or Unobtrusive Observation.

**BEHAVIOR CHANGE** interests you.

A **PARTICIPANT** is the source of data.

Then try: Self-Evaluation, Direct Observation, Games Observation, Participant Observation, or Unobtrusive Observation.

A **PARENT**, **VISITING TEACHER**, or **COMMUNITY RESOURCE** is the source of data.

Then try: Personal Interview, Anecdotal Recording, Focus Group, or Written Questionnaire.
If a **Program Objective** is your evaluation object.

**Behavior Change** interests you,

A STAFF member is the source of data.

Then try:  
- Self-Evaluation,
- Personal Interview,
- Direct Observation,
- Unobtrusive Observation,
- Focus Group, or
- Written Questionnaire.

**Cognitive Change** is of interest,

Then try:  
- Test,
- Standardized Indicator,
- Artifact Analysis, or
- Self-Evaluation.

The charts on the following pages link techniques to purpose through the sources of data. Review them to help in your selection of techniques. Figures 12 is a graphic representation of the material.
FIGURE 12: Connecting Techniques to Purposes Through Sources of Data

The following charts link purposes (topics) for evaluation to appropriate techniques through sources of data. There is one chart for each of the objects of evaluation. Based on the purpose that you wrote on p. 30, locate the chart which suits your needs. Use the chart to identify techniques that you can use, based on the sources of data to which you have access.

| PROGRAM |

<table>
<thead>
<tr>
<th>TOPIC</th>
<th>SOURCE</th>
<th>TECHNIQUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual events</td>
<td>Outside Observer</td>
<td>Unobtrusive Observation</td>
</tr>
<tr>
<td></td>
<td>Staff</td>
<td>Direct Observation</td>
</tr>
<tr>
<td></td>
<td>Visiting Teacher</td>
<td>Participant Observation</td>
</tr>
<tr>
<td></td>
<td>Parent</td>
<td>Check Sheet</td>
</tr>
<tr>
<td></td>
<td>Participant</td>
<td>Journal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Anecdotal Recording</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Experience Sampling</td>
</tr>
<tr>
<td>Behaviors</td>
<td>Outside Observer</td>
<td>Direct Observation</td>
</tr>
<tr>
<td></td>
<td>Staff</td>
<td>Participant Observation</td>
</tr>
<tr>
<td></td>
<td>Participant</td>
<td>Unobtrusive Observation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self Evaluation</td>
</tr>
<tr>
<td>Thoughts and attitudes</td>
<td>Staff</td>
<td>Artifact Analysis</td>
</tr>
<tr>
<td></td>
<td>Participant</td>
<td>Self Evaluation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Written Questionnaire</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Participant Observation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Personal Interview</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Experience sampling</td>
</tr>
</tbody>
</table>
TECHNIQUES

PROGRAM PURPOSE

CRITERIA

SOURCE

TECHNIQUE

External

Outside Expert
Principal
Visiting Teacher
Parent
Participant

Outside Consultant
Q-Sort
Written Questionnaire
Focus Group
Personal Interview

Internal

Program Administrator
Board of Directors

FACILITIES

TOPIC

SOURCE

TECHNIQUE

Satisfaction

Parents
Participant

Written Questionnaire
Interview
Focus Group
Check Sheet

Safety

Visiting Teacher

Efficiency

Outside Observer
Staff

Accessibility

Certifying Organization
TECHNIQUES

STAFF

TOPIC

Performance

SOURCE

Participant

Peer

Visiting Teacher

Outside Observer

Program Administrator

Staff

TECHNIQUE

Written Questionnaire

Interview

Anecdotal record

Direct Observation

Self Evaluation

Program Administrator

Organisational design

Board of Directors

Focus Group

Written Questionnaire
TECHNIQUES

SPECIFIC PROGRAM OBJECTIVES

TOPIC

SOURCE

TECHNIQUE

Experience Sampling
Sociometric Technique
Self Evaluation
Artifact Analysis
Standardized Indicator
Interview
Written Questionnaire
Focus Group
Anecdotal Record
Participant Observation
Unobtrusive Observation

Participant
Parents
Visiting Teachers
Community Resources
Program Staff
Outside Observer

Participant
Parents
Visiting Teacher
Community Resource
Program Staff
Outside Observer

Participant

Self Evaluation
Interview
Anecdotal Recording
Behavior Observation
Written Questionnaire
Direct Observation
Participant Observation
Unobtrusive Observation

Test
Standardized Indicator
Artifact Analysis
Self Evaluation

Affective Change in Participant, Parents, Visiting Teachers, or Staff

Behavior Change in Participant

Cognitive Change in Participant
Since the ultimate goal of many programs is to affect participants' behavior, an effective way to evaluate may be through observing the behavior of the participants. Behavior observation is often conducted for the evaluation purposes of: documenting a change in behavior, recording the occurrence of events, assessing the students' ability to apply knowledge or skills, or evaluating group interactions and dynamics. Types of behavior observations discussed in this guidebook are: (1) anecdotal records, (2) direct observation, (3) games observation, (4) participant observation, and (5) unobtrusive observation. Prior to a discussion of each of these, concise overviews are provided discussing: 'Why should I use them?’, 'Advantages', 'Disadvantages', and 'More information.'

**Why should I use them?**

Because experiential education (where learners are provided “hands-on” learning opportunities) is a keystone element of most programs, observation of the actual experience is an effective way to assess the quality of the program. Each of the behavior observation techniques provides a unique way to gain an in-depth understanding of the program. **Anecdotal records** are informal collections of comments, letters, or stories, relating what happened, how participants acted, and how they were affected by the program. **Direct observation** involves a non participant who watches and records behaviors during the program. **Games observation** involves the participants in a particular type of game in which desired actions or the lack of these actions can be observed. **Participant observation** requires the observer to become a participant in the program and then to record observations from an insider's point of view. **Unobtrusive observation** techniques are used to collect data about participants' behaviors without interacting with the participants (an example would be counting the litter left at a camp site after campers have left).

To use these techniques, you can observe specific behaviors in a contrived situation or as events naturally occur. If you control the setting (i.e., you set up a situation in which a participant has a chance to litter) you can force the participants into choices about specific actions (in this case littering). A natural setting will allow you to observe only spontaneous actions or interactions occurring within the program. If your evaluation is focused on a particular type of behavior (aggressiveness, leadership, or group cooperation), contriving a setting may be very effective. If the purpose of your evaluation is to determine what happens in the program, observation within the natural setting will be more appropriate.

**Advantages (General)**

- Is excellent for determining behavioral changes and for describing the program.
- Determines actual behaviors as opposed to intended or reported ones. (For instance, leaders may report that they never raise their voices when actually they do.) These techniques may help you to reduce error, especially if you are examining socially unacceptable behaviors.
- Does not interrupt the program and requires no effort on the participant's part.
- Is low-cost and may be informal.
Disadvantages (General)

- Observer bias and missed data. Since we are constantly bombarded with information, each of us has unconsciously developed selective perception. Even a well-trained observer will not notice everything that is happening and may miss important information. To support the validity of your findings, use multiple observers, video tape, and re-observe the events. Have some of the participants read and comment on the accuracy of the final report. You can check the reliability with a test-retest method. (See Appendix 1 for a description of the test-retest method.) If you make a videotape of some actions and ask the observers to view the tape and to make a record of the behaviors at two different times, then you can compare their observations for consistency.

- Interpretation of actions is difficult. You may run into questions such as: What does this action mean, and why did they do it? Actions do not always reveal the intention behind them. Caution must be taken not to infer attitudes and intentions from observed actions.

- Unrealistic to make a large number of formal observations.

- Hawthorne effects. People may act differently when they know that they are being observed. To diminish these effects, have the observer stay with the group for extended periods of time in the hope that the participants will return to their "normal" behaviors. You can use a hidden video camera to provide data that are not threatened by the observer's presence. (These also allow for in-depth and repeated analysis but must adhere to ethical principles such as right s of privacy.) To eliminate Hawthorne effects, you can use unobtrusive or participant observation in which the observer is unknown as an evaluator to the other participants.

- Time. All of the behavior observation techniques require large quantities of time for on-site observing and for data analysis.

- Data analysis. You can record the data as numerical frequencies of specified behaviors or as detailed written descriptions of detailed observations. In the first case, you may run into difficulty in identifying, defining, and categorizing particular behaviors, but once you have completed this, the analysis of the data should be straightforward. Descriptions allow you to relate all of the variables which may be influencing the actions, but analysis of descriptions may be confusing.

More Information (General)

Defining Behaviors:


Easily tailored to and designed for unique settings, behavior observation can be a versatile way to document the actual behavior changes of participants.
Technique: Anecdotal Records

Anecdotal records are informal collections of verbal comments, letters, stories, guest book comments, informal observations, or descriptions about the program. These can provide background information and examples of events within the program, but they are not always representative of everyone's opinions. Since these often come from secondary data sources (i.e., parents, staff, or visiting teachers), caution must be used in their interpretation. From studying collections of these records, you might discover a trend of recurring complaints or compliments which might provide insight into long-term problems with or benefits of the program.

Example

Dear Camp Bravo,

Thank you for the wonderful opportunity that you provided my daughter Alicia. She had such a good time and has come back a new person. I am amazed at how much more self-confident she is and how she now takes on new challenges by herself. Your program and staff should be highly commended for the excellent job! We are looking forward to another exciting summer next year!

Thank you again,

Mrs. Myers

Advantages
+ Easy.
+ Inexpensive.
+ Source of data may span a long period of time.

Disadvantages
- Not reliable.
- Unrepresentative of all opinions.
- Heavily weighted toward the positive.

More Information (See p. 73, the general section on behavior observation.)

Anecdotal records are an accessible and inexpensive source of information for an evaluation. However, since the information may have some bias, this technique is best for use in conjunction with other techniques.
Technique: Direct Observation

Direct observation is a popular technique for collecting data about behaviors. The most common formal approach for recording behaviors is for a trained observer to use a rating sheet to record specific behaviors as they occur. The rating sheet provides a mechanism for recording the occurrence and the frequency of any pre-specified behavior. Depending upon the evaluation purpose, behaviors of one individual or those of a group may be recorded.

This technique is probably most often used to study the interactions between individuals or at therapeutic programs (medical, behavior modification, etc.) to help determine participants’ progress towards individualized goals.

Advantages

+ Has diverse uses: documenting personal interactions, group dynamics, leadership abilities, cooperation, social maturity, expressions of attitudes, choice of behaviors, and ability to apply knowledge.
+ Records actual behaviors. It is a particularly useful tool to verify perceived or intended behaviors that were reported in another stage of the evaluation.
+ Is not limited to on-site use. By employing teachers, parents or others in the outside community (Boy/Girl Scout leaders, social workers, friends) as observers, this technique may be extended to gather information on the participants' behaviors when they are not at your site.
+ Is low-cost.

Disadvantages

- May be difficult to define or interpret specific behaviors. (For example, one child hits another. Was it an aggressive or a playful act?) You can limit these uncertainties by providing in-depth training for observers, by writing a detailed description of the event, setting, and conditions, and by not assuming a persons' intentions from his/her actions.
- Requires training for observers.
- Demands extensive on-site observation time and a low observer to participant ratio.
- Is susceptible to observer bias.

Should I make my own rating sheet?

Several standardized observation rating sheets have been developed (see Figure 13). For example, Warner (1984) created an observation check sheet to record levels of an individual's involvement with a group in completing several initiative tasks. This was used to demonstrate that students who had completed an experiential education trip demonstrated significantly higher levels of positive involvement than did students in a control group. An advantage of these standardized sheets is that they are already designed and have been tested for reliability and validity. You may want to develop your own scale so that you can pin-point specific behaviors of interest to you. Rating sheets which are developed by the evaluation team may provide much practical information, but they require testing to confirm their validity and reliability.
FIGURE 13. Standardized Behavior Rating Forms

<table>
<thead>
<tr>
<th>TOPIC</th>
<th>STANDARDIZED FORM</th>
<th>REFERENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biracial interactions</td>
<td>Self-developed</td>
<td>Martin (1975)</td>
</tr>
<tr>
<td>Classroom behavior</td>
<td>Hahnemann Elementary School Behavior Rating Scale</td>
<td>Spivack &amp; Swift (1975)</td>
</tr>
<tr>
<td>Competence</td>
<td>Self-developed</td>
<td>Ulrey (1974)</td>
</tr>
<tr>
<td></td>
<td>Cheffers' Adaption of the Flander's Interaction Analysis System</td>
<td>Robb (1981)</td>
</tr>
<tr>
<td></td>
<td>Self-developed</td>
<td>Roland (1983)</td>
</tr>
<tr>
<td>Leadership</td>
<td>Self-developed</td>
<td>Ulrey (1974)</td>
</tr>
<tr>
<td>Likert-type scale</td>
<td>Self-developed</td>
<td>Oakley (1980)</td>
</tr>
<tr>
<td>Participant involvement</td>
<td>Individual Response Gestalt II</td>
<td>Robb (1981)</td>
</tr>
<tr>
<td></td>
<td>Self-developed</td>
<td>Warner (1984)</td>
</tr>
<tr>
<td>Therapeutic behavioral goals</td>
<td>Goal Attainment Scale</td>
<td>Freeman, Anderson, Kairey &amp; Hunt (1982)</td>
</tr>
</tbody>
</table>

More Information (See p. 73, the general section on behavior observation.)

Self-made Scales:


Direct observation is a valuable method for recording the actual processes that occur in an outdoor education experience.
Technique: Games Observation

Games observation requires observing individuals while they are playing a certain type of game. As the game controls the setting, it may be used to encourage actions and interactions to be evaluated. This is particularly useful for exploring a group's social interactions. For example, an evaluator could view and record the actions of persons playing the "Chocolate Game" in which each person must make a choice either to cooperate with their peers or not (called individualistic). The pattern of choices determines the amount of chocolate that each individual will ultimately receive. The total group reward is not set in the beginning, but is dependent on the group's choices. The greater the number of cooperative choices, the higher the total reward to the group. If a person makes an individualistic choice, however, that person will receive lots of chocolate relative to the others who were cooperative. If a person makes a cooperative choice and others do not, then the individual will receive only a modest reward. This is an unique way to study the cooperation and trust among members of a group. (See Warner, 1982 for more details.)

Through watching individual and group actions and reactions, the evaluator could record information on the participants' behaviors and attitudes. Recording of data may be facilitated by the use of a video camera and a behavior rating form.

The technique, games observation, can be used with other games such as initiative activities (Peanut Butter Pit, The Wall, Bear Claw; see Cow Tails and Cobras for descriptions of these and other such games), and role playing. These activities can be used to determine pre/post program group cooperation, social patterns, and problem-solving abilities. As an example, consider how an environmental education program might use role playing to determine if students could apply acquired knowledge. The instructor could give the students a brief description of a dilemma (i.e., a town government trying to decide whether or not to spray pesticide in their town). Each student would assume the role of one of the characters (mayor, parents, scientists, pesticide manufacturer, owners of a golf course) and would then decide how that person would react. This might require research and discussion. Eventually, the students would hold a town council meeting to debate the issue and to make a final decision. Observers could watch the debate in order to determine how well the students understand the basic concepts and issues involved.

Advantages
+ Provides an experiential learning activity appropriate for outdoor education.
+ Is enjoyable for the participants.

Disadvantages
- Is susceptible to the effect of other variables. Close attention must be paid to the conditions under which the games are being played. Other factors such as how hungry or tired the group is may have a dramatic effect on their behaviors.
- May not be taken seriously by the participants.
- May be limited in the types of variables that can be studied.

More Information (See p. 73, the general section on behavior observation, and p. 76, the section on direct observation.)

Games:


**Games observation is a creative and fun way to conduct a study of a controlled situation in order to assess particular behaviors.**
Technique: Participant Observation

This technique is often used by those who feel that a non-participating observer cannot fully comprehend the experiences of those participating in the program. What better way to study the participant's attitudes, behaviors, or social interactions than by becoming a participant yourself? This process has been heralded as a way to study complex social interactions through an intensive observation of actions in the context in which they are performed. Usually, the participant observer will record observations in a hidden manner (often using a log book), attempting to create a summary of the many details observed in the everyday life of the program. Then, reviewing the notes, the observer may see patterns emerging from the observations and may create a framework for organizing the interpretations of the data.

Advantages

+ Data directly rooted in personal experiences. As a result of the intense, experiential study, participant observation can yield detailed contextual data on multiple topics.
+ Face-to-face interaction. This gives the observer the ability to probe for more information and to clarify confusing issues.
+ Little threat from Hawthorne effects.
+ Discovery of obscure information such as relationships, nonverbal communication, unconscious thoughts, and behaviors.
+ Low cost. Typically participant observation requires only paper and pencil.

Disadvantages

- Time-consuming.
- Requires a perceptive and devoted observer. Observer must be able to conceptualize general patterns in the data while not losing sight of the minute details of a situation.
- Observer bias. The entire study is based on one individual's observations and interpretations. To support the validity and reliability of the results, observers should: (1) write in-depth descriptions of their own actions and the procedures that were followed; (2) search for alternative explanations; (3) demonstrate that conclusions which oppose theirs cannot be confirmed; and, (4) ask participants of the program to review the final report in order to verify its accuracy.
- Missed data and misinterpretation.
- Ethical dilemmas may result if the other participants are not aware that their actions are being observed and recorded.
- Difficulty in replication (Henderson, 1991).

More Information  (See p. 73, the general section on behavior observations.)

General:

**BEHAVIOR OBSERVATIONS**


Examples:


---

_Time and devotion of the observer are the crucial elements for a successful evaluation through participant observation._
Technique: Unobtrusive Observation

These are indirect observations of people and their actions. Often this is accomplished through gathering physical evidence of peoples' actions after they have left a site (i.e., how much food was wasted in the dining hall, through electronic recording devices (i.e., hidden cameras), or through an inconspicuous observer noting the type of activity and conversations. An unobtrusive observer is like an undercover detective. The evaluator seeks information about the program in a covert manner so that the observer's presence will not affect the participant's behaviors.

Observations can be conducted in natural or contrived situations. For example, to determine a program's effect on campers' concern over litter, evaluators could use either of the following procedures. They could count the pieces of litter left in an area after a snack, (a natural situation), or they could plant litter along a trail that the campers use regularly and then count how much was picked up (a contrived situation). As long as the participants have no knowledge that their actions are being recorded, both of these would be unobtrusive measures.

Some examples
- Noting peoples' choice of recreational activities.
- Observing the neatness and cleanliness of individuals' bunk areas.
- Recording voluntary attendance at an environmental organization.
- Counting the number of books checked out from the library on the topic of interest.

Advantages
- Eliminates Hawthorne effects. There is no influence of the observer on the participants.
- Captures socially unacceptable behaviors.
- Assures anonymity.
- Does not interfere with the program.
- Places no demands on the participant; is not time-consuming for participants.
- Is low cost.
- May not require long periods of observation.
- Records actual behaviors; is an excellent mechanism for verifying intentions to behave that were reported through another technique.

Disadvantages
- Little control over the situation. Thus, the influence of extraneous variables is unknown and care must be taken in the interpretation of the data.
- Dependence on inference. To make conclusions you might have to infer attitudes or intended behaviors from actions. If someone throws a paper cup at a trash basket and misses, is that person is littering, did the individual intend to litter, did the person even know that they missed the trash basket, and would that make a difference? What conclusions could you make?
- Ethical concerns (e.g., observation in sleeping cabins might infringe upon individuals' privacy).

- Sampling error. If observations are limited to one location, there is a chance that the observations will not be representative of all of the participants' behaviors (i.e., those who do not go to the area under observation will never be observed).

- Often provides only assessment of the group as a whole. It does not usually focus on the behaviors of individuals.

More Information [See p. 73, the general section on behavior observations.]


Unobtrusive observation measures are a relatively easy method for gaining information about actual behaviors.
WRITTEN QUESTIONNAIRES (General Information)

Written questionnaires are a familiar technique for collecting information. Questionnaires are easy to administer, are productive, and can be designed for almost any evaluation purpose. Unfortunately, the flexibility of this method has led some evaluators to apply "the Rule of the Hammer" with it. These evaluators use written questionnaires for every evaluation without considering other more appropriate methods or even tailoring a questionnaire to suit the needs of their program and audience. This results in wasted time and energy and negative attitudes toward evaluations.

Advantages (General)

+ Ability to design your own specific questionnaire.
+ Access to stakeholders who are not on-site through mailed questionnaires. By using this as a follow-up after participants have left the program, you can gather information on how long the program effects last.
+ Immediate or delayed response. This gives the respondent time to think about complex issues before responding.
+ Flexibility. Questionnaires can be created to collect unique types of data from a wide variety of stakeholders.
+ Low cost if conducted on-site and analyzed by staff. For mailed questionnaires, postage, copying costs, and follow-up contacts must be considered.
+ Adaptable time schedule. Questionnaires can be distributed in a compact unit that combines directions, questions, and data recording, completed independent of the evaluator. This reduction of personal contact reduces the time demand typically required by the evaluator. Time is spent mostly in preparation (especially if you are writing your own questionnaire) and in the interpretation of open-ended questions.

Disadvantages (General)

- Non-Response. This is the greatest problem with written questionnaires, particularly with mailed ones. It is easy for people to not answer either particular questions or the entire questionnaire. When the non-respondents, as a group, differ from the respondents, non-response error can cause tremendous error in your conclusions. To better understand this, consider the following scenario: In planning for an adventure education program for troubled children, the program staff asked some of the children’s social workers for help in planning activities. Other social workers who worked with children who would be coming to the camp were not included in the planning stage. After the program, questionnaires were sent to all of the social workers. The social workers who had been included in the planning stage had a personal interest in the program’s design and its success; therefore they readily responded to the survey. The others felt that they were left out and thus had a negative attitude toward the adventure education program. When these social workers received the questionnaire, they refused to answer it or provided only negative feedback. Obviously, those who were left out had something valuable to say. If the non-respondents were ignored, valuable data would be lost. (See Appendices 4 and 6 for ways to reduce non-response.)
Typically provide information on perceptions and intentions to act. Because questionnaires rely on the respondents' retelling of facts and because questionnaires are often used with a single case study approach, many researchers feel that questionnaires are inferior to other techniques. Yet, this collection technique can provide meaningful results and can be used with experimental designs such as pre/posttest and between-group comparisons.

Language misunderstandings. Using written questions causes problems for those who cannot read or comprehend the language of the questionnaire. Additionally, poor phrasing of questions may result in differing interpretations by the respondents. Questionnaires require careful construction and pretesting to ensure that they are easily understood. (See Appendices 1 and 6.)

Demands on the respondents. This technique relies on respondents' honesty. If you are surveying program participants while they are on-site, it will also cause an interruption in the program. Consider using slides or pictures in conjunction with a written questionnaire to stimulate the interest of the respondents. For example, if you are interested in participants' opinions of certain activities, you could show a slide of each activity to an entire group of participants and ask each participant to respond to questions about the activity.

Hawthorne effects. Care must be taken to establish a non-threatening, unhurried atmosphere in which people are assured of anonymity and do not feel pressured to offer only socially desirable answers. (The last night of a summer camp is not an ideal time to record attitudes toward camp.)

More Information (General)


Written questionnaires can be especially useful for gathering information from individuals who are off-site. However, questionnaires require careful construction and testing to insure that they are easily understood.
Types of Questions

Many types of questions (items) can be used in questionnaires. Choose the type which is appropriate for your audience. The types described in detail here are: open-ended, scaled, Likert and Likert-type, Thurstone scale, Guttman scale, and semantic differential. In addition to these, you may also want to include some factual or closed-ended questions (i.e., for gathering basic information such as age, sex, date of visit, favorite activity).

Question Type: Open-ended

Open-ended questions allow the respondents to freely express their opinions, attitudes, and inputs while permitting the evaluator to frame the question.

Example: [For parents]

Describe any changes that you have noticed in your child's behavior since he/she has returned from the Outdoor School.

Advantages

* Discovery of unanticipated outcomes.
* Personal, in-depth and descriptive information.
* Non-constraining. By encouraging people to respond in their own words, you do not limit them to pre-established response categories. If a person has a genuine interest in providing valid data, they will probably prefer this freedom and will give thorough, thoughtful responses.

Disadvantages

* Large amount of effort required of respondent. Some people may only give short or superficial answers or may skip open-ended questions all together.
* Difficult and time-consuming analysis, resulting from the qualitative nature of the data.
* Not appropriate for young children or those who cannot read or write.
* Biased collection. Valid data may be gathered only from respondents who understand the question, have good writing skills, and are willing to express themselves.
* Lack of structure. A questionnaire consisting of only open-ended questions may not provide enough guidance or may be too taxing for respondents. Therefore, these questions must be carefully designed to encourage responses. For example, a broad question such as, "What did you like about your adventure vacation?", could be better structured as several more specific questions such as, "On your adventure vacation, what activities did you enjoy?" and "Describe what you liked the most about each of those activities."

* Potential bias by the evaluator. If questions are written in an inappropriate manner, respondents may be lead toward a particular answer.
More Information: (See p. 84, the general section on written questionnaires.)


Question Type: Scaled Items

Scaled items are used to gather information on how people rank order or rate certain activities, staff or program objectives. Notice the difference in the rank ordering scale (demonstrated in Example 1) and the rating scale (demonstrated in Example 2). Rank ordering forces the respondent to choose one item over another.

Example 1: [For board members]

Example 2: [For staff of a program for people who use wheelchairs]

On a scale from 1 to 10 (1 being the strongest), rate the following activities that you conduct on levels of required agility, strength, and muscle coordination.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Agility</th>
<th>Strength</th>
<th>Coordination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensory awareness hike</td>
<td>_____</td>
<td>_____</td>
<td>_____</td>
</tr>
<tr>
<td>Cabin Clean-up</td>
<td>_____</td>
<td>_____</td>
<td>_____</td>
</tr>
<tr>
<td>Cook-out</td>
<td>_____</td>
<td>_____</td>
<td>_____</td>
</tr>
</tbody>
</table>
Advantages
+ Ideal for gathering information about priorities.
+ Easy analysis of quantitative data.

Disadvantages
- Results lack detail.
- Boring for the respondent.
- Confusing for the respondent. Respondents may find it difficult to rank order more than three items and may not understand rating scales.

More Information (See p. 84, the general section on written questionnaires.)

Examples:

Question Type: Likert and Likert-type

Likert and Likert-type questions (named for R. Likert who developed the scales) ask participants to identify the degree to which they agree or disagree with a statement. Typically a five or seven point scale is used to quantify the individual's feelings toward a non-neutral statement about the object. Likert-type scales use the same format but different descriptors, such as: importance, frequency, and likelihood.

One type of evaluation that uses Likert-type scales is the Importance/Performance analysis. This type of evaluation asks respondents to score parts of the program (i.e. overnight trips, craft time, etc.) in terms of how important they are to the individual and how well the program performed (or the program’s quality). This can be accomplished through double-scaling (see Example 4) or through separate measures. Double-scaling, in which two scales are presented on opposite sides of one item, is a time and space saving device. However, because of the close proximity of the scales, there is some concern that the respondents will not distinguish between the importance and performance attributes and will unintentionally score them identically. (See the section on Importance/Performance analysis on page 44 for more information.)
Example 1: Likert [For a visiting teacher]

Using the response categories listed below, note your opinion on the following statements:

1 = strongly agree, 2 = agree, 3 = neutral, 4 = disagree, 5 = strongly disagree.

1. The staff at The Field School are knowledgeable in their subject area.
2. The staff at The Field School are not friendly.

Example 2: Likert-type [For a visiting teacher]

Consider the teaching techniques that you have learned at the Outdoor School. Since your visit, how often have you used the following with your class (please use a check mark to indicate your response):

<table>
<thead>
<tr>
<th></th>
<th>NEVER</th>
<th>SELDOM</th>
<th>SOMETIMES</th>
<th>OFTEN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(once every six months)</td>
<td>(once a month)</td>
<td>(once a week)</td>
<td></td>
</tr>
<tr>
<td>Discovery learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensory awareness activities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Example 3: Likert-type [For a young participant]

Circle the face that shows how you felt when you were in the woods.

<table>
<thead>
<tr>
<th></th>
<th>VERY UNHAPPY</th>
<th>UNHAPPY</th>
<th>O.K</th>
<th>HAPPY</th>
<th>VERY HAPPY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&quot;[sad face]&quot;</td>
<td>&quot;[unhappy face]&quot;</td>
<td>&quot;[neutral face]&quot;</td>
<td>&quot;[happy face]&quot;</td>
<td>&quot;[very happy face]&quot;</td>
</tr>
</tbody>
</table>
Example 4: Likert-type (For staff)

On a scale of one to seven, please rate how important each of the following program elements is to you and rate the quality of each of the elements. The higher the number, the higher the importance and the better the performance.

<table>
<thead>
<tr>
<th>IMPORTANCE</th>
<th>QUALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Advantages
+ Is easy to complete.
+ Is easy to score.
+ Can be designed for a non-reader. Consider reading the questions to the respondents and including pictures on the scale. (See example 3 above.)
+ Allows for flexibility. You can force the respondent to chose a positive or a negative response by not including a neutral option (i.e. providing an even number of response categories). Or, you can allow neutral responses with an odd number of response categories.

Disadvantages
- Assumption: an attitude can be placed on a continuum from completely negative to completely positive.
- Statements must be positively or negatively phrased. This is subject to individual interpretation, meaning that construction of such a scale requires thorough checks on validity and reliability.
- No absolute zero or neutral score. Scores can only be interpreted as relative to each other and are not comparable to scores from other questionnaires.
- Response sets. The ease of response (i.e., simply checking a box) may lure a person into not considering each item individually and thus answering haphazardly (i.e., going down the page and checking all “strongly agree” just to complete the questionnaire).

More Information (See p. 84, the general section on written questionnaires.)

Examples:
Question Type: Thurstone Scale

Thurstone scales ask the respondents to mark either their agreement or their disagreement to each of a list of eleven statements about an object. Each statement represents a point an equal distance from the others along the spectrum of possible attitudes. The statements are assigned weights according to their position on the continuum (i.e., the most negative statement would be weighted a one, the neutral statement would be weighted a six, and the most positive would be weighted an eleven). The weights of the items that were agreed with are averaged to compute the respondent's score. The example below further demonstrates the scoring of these items.

Example:
Please place an X by each statement with which you agree.

Advantages

<table>
<thead>
<tr>
<th>Number</th>
<th>Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>It is absolutely ridiculous to be concerned for the health of our planet.</td>
</tr>
<tr>
<td>8</td>
<td>People should not litter.</td>
</tr>
<tr>
<td>11</td>
<td>I need to devote my life to saving the planet which is in a state of disaster.</td>
</tr>
<tr>
<td>4</td>
<td>Environmental concerns are not my business.</td>
</tr>
<tr>
<td>5</td>
<td>Environmentalists can do what they want as long as they don't interfere with me.</td>
</tr>
<tr>
<td>7</td>
<td>Recycling is a good idea.</td>
</tr>
<tr>
<td>2</td>
<td>Humans control the world and don't need to worry about polluting too much.</td>
</tr>
<tr>
<td>9</td>
<td>People should be more environmentally conscious in their choice of lifestyles.</td>
</tr>
<tr>
<td>3</td>
<td>If a species is endangered, it shows that they are weak, and they should die.</td>
</tr>
<tr>
<td>6</td>
<td>Pollution is not great, but it's not really hurting me.</td>
</tr>
<tr>
<td>10</td>
<td>We must stop pollution!</td>
</tr>
</tbody>
</table>

This respondent's score = \( \frac{8+6+9}{3} = 8 \)

Disadvantages

- Assumption: Attitudes can be placed on a continuous spectrum from positive to negative.
- Complexity.

Construction is extremely complex and requires identification of statements that represent a balanced continuum of attitudes about the topic.

More Information (See p. 84, the general section on written questionnaires.)

Question Type: Guttman Scale

The Guttman scales ask individuals to read a set of statements and to mark those with which they agree. This scale is built on the principle that attitudes can be placed on a continuum. Attitudes are accepted as unidimensional (i.e., maybe your attitude about hunting can accurately be placed on a scale from good to bad).

```
Hunting is:
GOOD---------------X----------------BAD
```

The scale consists of statements that represent attitudes along the continuum. In the example below, if a respondent agrees with statement C, then logically the respondent could be expected to agree with statements D and E but would disagree with statements A and B. Guttman scales require that the respondent totally agrees with statements which are listed below the one which is selected.

Example:

Please place an X beside all of the statements with which you agree:

```
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
</tr>
<tr>
<td>X</td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>E</td>
</tr>
</tbody>
</table>
```

Advantages

+ Precise results.
+ Relatively easy to complete.

Disadvantages

- Requires that the attitude is defined narrowly.
- Assumes that attitude is unidimensional. This may be a big assumption. For example, a person's attitude about hunting may be more complex than acceptable/unacceptable.
- Difficult to construct.

More Information  (See p. 84, the general section on written questionnaires.)

Question Type: Semantic Differential

The semantic differential is a popular technique that uses pairs of contrasting adjectives (good/bad, hot/cold) to describe an object, person, activity, etc. Semantic differentials can be designed to determine the individual's opinion about separate characteristics of the object (i.e., each pair of adjectives counts as one question) or the overall opinion of the object (in which case, the responses to all of the adjectives within a set would be averaged to measure the attitude). A creative adaptation of this method is to show respondents pictures or slides instead of, or in addition to, the written topic.

Example: [For a parent]

Please consider the center's physical facilities. Under each of the topics below is a list of scales. Please rate the facilities by placing an X in a space nearest to the adjective that best describes the facility.

<table>
<thead>
<tr>
<th>Clean</th>
<th>Spacious</th>
<th>Uninviting</th>
<th>Easy to use</th>
<th>Inaccessible</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

This parent felt that the dining hall was moderately clean, very confined, neither inviting nor uninviting, moderately hard to use and very inaccessible. Note that not all of the positive descriptors are on the left. This forces individuals to consider each pair of adjectives separately and reduces the likelihood of response sets.

Advantages

+ Easy to complete.
+ Easy to design, administer, and score. Lists of opposing adjectives are available.
+ Obvious "good" and "bad" responses. This straightforward approach appeals to many people who may feel that they are being tricked with some of the other techniques.

Disadvantages

- Face validity. Respondents may be confused by adjectives that do not seem to directly describe the object. (An example of this is the use of the sharp to dull scale for rating a person.) In such cases, the respondent may skip the item. To eliminate this problem, choose only descriptors that relate to the object.
- Obvious "good" and "bad" responses. This may cause problems if individuals consciously try to slant their responses.
- Interpretation of words. For example, "accessible" could refer to ADA standards, having a friendly or open manner, or having a central location.
More Information  (See p. 84, the general section on written questionnaires.)

General:


Examples:


VERBAL TECHNIQUES (General Information)

When you ask people to voice their opinions, you give them a greater role in the evaluation process, you empower them, and you encourage a more positive attitude toward the evaluation. The most common verbal techniques include: focus groups and telephone or personal (face-to-face) interviews. These methods are easily adapted to unique situations, can be used in conjunction with other techniques, and often are used as follow-up procedures. Before the individual techniques are described, the general issues of: "How much structure?", "Advantages", "Disadvantages", and "More information" are discussed in the following sections.

How much structure?

The amount of structure in the verbal techniques varies widely, ranging from formal, structured interviews to spontaneous, informal conversations. Depending upon the structure and type of questions asked, you can collect either qualitative (words and descriptions) or quantitative (numbers and statistical) data. If you use a structured list of closed-ended questions (where the person must choose from a list of answers), then responses can be easily recorded on a standardized form, tallied, and analyzed. This ease comes with a price, however. The list of questions and answers (called an interview schedule) may be so structured that it is limiting. In some cases the respondent may not agree with any of the responses provided, or may want to clarify or elaborate on a response, but is not given the opportunity.

The use of a standard list of open-ended questions allows respondents choose their own words in responding while still providing some structure for the conversation. You may opt to provide only an outline of the topics that you want each respondent to address. This gives even more flexibility, but it requires an evaluator who can guide the conversation toward, and keep the respondents focused upon, the specific topics of interest. For total flexibility, you can use informal conversations with no prescribed structure. This technique can be very valuable if the evaluator is perceptive and can draw patterns and meaning out of everyday conversations.

With the less structured formats, recording the responses may prove difficult. You will probably rely on note-taking to record the conversation; however, it is important that this does not interfere with the non-committal and free-flowing atmosphere of the discussion. If you have the respondents permission, tape record the communication for later analysis. In some cases, people may be particularly reluctant to respond if you are recording their responses.

The formal and structured approaches are easily analyzed and reported with traditional techniques (i.e., summing responses to individual questions and tabulating frequencies, percentages, and means). However, if there is little structure in the data, other approaches may be needed. If you use tape or video recordings, then you can transcribe the participants' comments word for word. Transcriptions are extremely time-consuming, but provides detailed information. The transcriptions can then be analyzed through content analysis. (See Appendix 5 for more on content analysis.)
Advantages (General)

+ Rich, in-depth information.
+ Ability to probe. Live interviewers and focus group facilitators can probe respondents for an explanation of a vague or shallow response, and they can clarify questions for the respondent. If an evaluator develops a friendly rapport with the respondent and establishes a relaxed atmosphere, the respondent will give frank, thoughtful, and honest information.
+ No reading required.
+ Pleasant for respondent. People enjoy talking, and verbally they can provide a great amount of immediate feedback. This may be a great advantage if you are asking respondents to provide information over an extended period of time, as they may tire of repeated written questionnaires.
+ Unique, personal, detailed, and qualitative data.

Disadvantages (General)

- Time. These techniques require a tremendous amount of time from the participants and the evaluator.
- Program interruption.
- Superficial responses. In face-to-face communication, respondents may feel pressured to create a quick answer which may not be thoughtful.
- Responsibilities of evaluator. The evaluator must create a conducive atmosphere, encourage participants’ responses, keep the conversation on the topic of interest, and phrase questions in an unbiased manner.
- Establishment of validity and reliability is necessary and may require extra efforts.

More Information (General)


The verbal data collection techniques are flexible, low-cost, and rewarding, but they are labor-intensive, ways to gather in-depth information for a variety of evaluative purposes.
Technique: Focus Group Interview (FGI's)

"I remember when I was young, every time my parents met with my friends' parents, they always talked about the activities their children were involved in. After our youth group came home from church camp, the popular topic would be how many clothes had been lost, how we had changed, and what things the camp should do differently. When my mother mentioned some of the positive effects of the camp, my friends' mother would want to tell the story of how her child was not scared of the dark anymore. From listening to them you would have thought that they were experts on the subject! After all, who would know more about the changes that were occurring in their children?" If you were the director of that camp, would you not have loved to be a fly on the wall during those discussions? A focus group that sets up an informal group discussion can give you similar information.

A focus group usually consists of at least one moderator (a person who facilitates and observes the discussion and records data) and a group of 8-10 stakeholders who discuss an issue(s) of concern to the evaluation. This method can be used to assess parents', campers', or teachers' opinions on the strengths and weaknesses of a program, to view initial responses to a proposed program or change in the existing program, to provide suggestions for future planning, or to explore the perceived benefits of a program (useful information for marketing purposes). The following example provides an outline of the steps in conducting an FGI.

Example:

I. Decide the purpose for the FGI
II. Develop a questioning route (an outline of open-ended questions and probes to be used)
III. Train facilitators and assistants
IV. Select participants who have some common characteristics (e.g., seniors who have all participated in a joint seniors and youth session)
V. Conduct FGI
   A. Provide a comfortable atmosphere
   B. Welcome participants and familiarize them with the study and its purposes
   C. Facilitate, observe, and record the discussion
VI. Analyze the data (adapted from Wells in Higginbotham & Cox, 1979)
   A. Review notes immediately and check recording devices
   B. Choose one of the following strategies for analysis:
      1. For a "quick and dirty" analysis, have the moderator(s) record their impressions. Use this only when limited time is a critical factor because it may result in missed or biased data.
      2. When time is limited or transcription of tapes is impossible, the moderator can listen to the tapes and pick out phrases which are relevant to support conclusions about prevailing themes in the data.
      3. For a detailed analysis, content analyze the transcripts.
      4. If multiple moderators can listen to tapes, discuss trends in the data, and ground their conclusions in actual quotations, then this type of integrated approach is highly recommended.
   Although time-consuming, this adds great validity and can result in a concise, informative synopsis of the interview.
V. Report data, make judgments, and apply the results.


VERBAL TECHNIQUES

Advantages

+ Social orientation. For FGI participants, this is enjoyable, gives them a feeling of empowerment, and stimulates them to think.
+ More efficient than personal interviews.
+ Allows for debate. The group can discuss issues, come to a consensus, and support their conclusions with personal examples.

Disadvantages

- Requirement of trained facilitators who can keep the discussion flowing yet focused.
- Difficult analysis.
- Time-consuming for both parties.
- Inconvenience of gathering stakeholders together. Sometimes this may require paying the participants.
- Hesitancy of participants to speak up against opinions of others.

More Information (See p. 95, the general section on verbal techniques.)

General:


Example:


A focus group can be thought of as a group interview which allows participants to share ideas with each other while the moderator facilitates, observes, and records.
Technique: Personal (Face-to-Face) Interview

To determine how people feel about your program, you may want to ask them directly. Personal interviews are a strong technique for exploring individuals' opinions. This technique can also be used in a pre/posttest fashion for determining changes in attitude towards the program, the environment, peers, school, etc. A skilled educator can use an interview to measure someone's knowledge, problem-solving skills, or mastery of concepts. Additionally, interviews can be used to request reports from parents, teachers, or staff on their observations on a child's behavior before, during, and after the visit. Also, the two-way communication possible in an interview makes it a suitable choice for staff evaluations.

Although an interview appears to be an easy way to collect information, it is important (and may be difficult) to establish validity and reliability. (These terms are further explained in Appendix 1.) Interviewers often create an interview schedule which is the proposed list of questions and prompting words that they plan to use. If the interview schedule is reviewed by a panel of experts (staff, other camping professionals), the validity can be determined (they can determine if the questions ask what you are intending for them to ask). The questions can be tested on people who are similar to the actual interviewees in order to identify questions and expressions which are vague or confusing. Since replication of any of the verbal techniques is difficult, establishment of the reliability may be best accomplished through tape recording the process and later reviewing the responses. Since evaluators typically interview only a small sample of people, it is important that the sample adequately represents the array of people that are being studied. (Refer to Appendix 3 for more information on sampling.)

The interviewer

The key to obtaining useful information from an interview may be in the interviewers themselves. You must decide who will be the least threatening to the respondents: a member of the regular staff who knows the program and the participants well, or an outsider. To make this choice, consider the types of questions that will be asked. For example, if the questions relate to an individual staff member's performance, participants may be reluctant to answer honestly to a staff member interviewer who personally knows the individual who is being evaluated. The available budget and staff time also may play a part in this decision. The interviewer(s) should be sensitive, objective, good listeners, and well-trained. The greatest challenge for the interviewer is to encourage the interviewee to provide a personal perspective on the situation. This can be accomplished through reassuring the interviewee that all responses are confidential (and they need to be), by creating a non-threatening atmosphere, and by thoroughly explaining the purpose for the evaluation.

Advantages

+ Clarification and personalization of questions. The interviewer can make the questions meaningful and understandable to the respondent, resulting in more in-depth responses.

+ Telephone interviews. These provide access to stakeholders who are rarely on-site, and they work well as follow-up procedures.
VERBAL TECHNIQUES

Disadvantages

-Time. A large commitment of time is required by both parties for the actual interview. The time required for analysis of the responses may be greater than that for any of the other techniques. Of course, this is dependent on the amount of structure and the type of recording you use.

-Program interruption. Normally only a few people miss an activity.

-Bias of interviewer. In order to not influence the interviewee's responses, interviewers must carefully control their phrasing of questions and their body language.

More Information [See p.95, the general section on verbal techniques.]

General:


Examples:


Personal interviews are an effective method for discovering personal insights into the benefits of, needs for, and processes occurring within a program.
MISCELLANEOUS TECHNIQUES

The following techniques do not neatly fit into the categories of behavior observation, verbal techniques, or written questionnaires. They are listed in this section of miscellaneous techniques in alphabetical order.

Technique: Artifact Analysis

Artifacts that have been created by program participants or staff often offer valuable information about the individuals' values, attitudes, thinking skills, or knowledge. Some examples are works of art, personal journals, and academic projects. These are particularly useful for gaining insights about the personal growth of individuals.

Using artwork

Individual creative works, such as a clay piece, a photograph, a painting, a poem, or an entire portfolio (selected pieces chosen from an complete set of works) can be examined to discover values, attitudes, and recurring themes or expressions of the person.

Using journals

NOTES FROM A COUNSELOR'S JOURNAL

It is amazing the things I have learned. Today, Nikita, who is usually so ornery, did the nicest thing. When she saw Sonia crying because she was the only one who did not get any mail, Nikita went right up and shared her own mail with Sonia. Maybe those group building activities really are working.

Nikita has taught me how some people put on a tough outer shell when they really do care about others. I realized that, through my work here, I can help children break down those outer walls and reach out to support each other.

This excerpt provides insight into this counselor's experiences. Consider how your evaluation audience would react if presented findings written in this manner as opposed to being told that 85% of the counselors felt that they had experienced personal growth through their work at the camp? Quotations from personal journals, completed progressively throughout the program, can be extremely powerful evidence of staff and participant impressions, likes, dislikes, and reactions to the program, peers, or leaders, as well as evidence of personal growth. Entries
written while on a solo (an experience during which an individual is secluded in nature for an extended period of time) are often particularly revealing.

Spontaneous journal writings may furnish excellent expressions about the outdoor education experience. However, these are often very personal, and the participant may be hesitant to share them with the evaluator. To solve this dilemma, some evaluators have asked staff and campers to use a journal with special tear-out pages. These pages could contain open-ended questions, suggested themes to write about, or blank space for free expressions and could be regularly collected. An extension of the use of journals is for the purpose of staff evaluation through self-reflection. To do this, have staff members describe and record their own actions and the effects of these actions. These comments can be shared with a trusted administrator who can help them address their strengths and weaknesses.

Using academic projects

Academic projects best provide information about knowledge, mastery of concepts and thinking skills, but they may also yield some data on feelings, interests, and opinions. The range of possible projects is immense. Some examples are: leaf collections, descriptive paragraphs, field data collection records, natural history reports, short stories written from a plant's perspective, etc.

Advantages

+ Unique, fun, and a learning activity in itself.
+ Possibility for uncovering feelings, experiences, and both positive and negative unintended effects.
+ Meaningful and valid data. This depends on the development of a trusting relationship between the evaluator and the respondents.
+ Extended collection. It facilitates the recording of personal expressions over an extended time period. Valuable for formative evaluations.
+ Multiple entries. These support the reliability of the data and provide a mechanism for tracing individual student development through the stages of the learning process.

Disadvantages

- Lengthy and difficult analysis. (Accomplished through the use of rating sheets, content analysis, or other pre-established criteria).
- Demands confidentiality and permission to use the artifacts.
- Large time and energy requirement for both parties.
- May not be well-suited for program assessment. Portfolios are best when used for assessing individuals’ growth; however, group projects (murals, trail maintenance, etc.) may provide information on changes in behavior and group cohesiveness.
- Reliance on other skills. You must ensure that only the knowledge or skill of interest is being tested (as opposed to skill in writing).
MISCELLANEOUS TECHNIQUES

More Information

General:

Portfolios:

Examination of artifacts is a useful approach for obtaining progressive information about the growth of individuals.
Technique: Check Sheet

A check sheet can be completed by one or several observers as an efficient and effective method for assessment. Although this technique may be applicable to other program aspects, check sheets are most often used to rate a program's facilities, site, or balance of activities. Check sheets can be used by any objective reporter, including, administrators, staff, and outside consultants. A check sheet may be a simple check list, recording only the presence or absence of certain objects, or may be a rating scale similar to those on written questionnaires. Check sheets differ from written questionnaires in that they do not measure attitudes or opinions; rather, they address the perceived state of the program.

Advantages

+ Efficient for accreditation purposes. They are frequently used by organizations such as the American Camping Association in accreditation visits.
+ Ease in design, use, and analysis.
+ Inexpensive. The major expense is the recorders' salary.
+ Quick.

Disadvantages

- Reliance on observers' objectivity.
- Not usually appropriate for gathering process or output data.

Some uses of check sheets are for: judging the cleanliness of buildings, rating the accessibility of buildings to all persons, recording the level of nutrition provided, identifying safety hazards, or noting the structure of the activities provided.

More Information

Accessibility:


General Facilities:


Program Balance:


Check sheets are a most efficient methods for recording data about program inputs or facilities.
Technique: Experience Sampling

Experience sampling is an innovative technique currently in use in the field of recreation and leisure research to sample peoples' emotional states at random points throughout their outdoor experience. For example, participants might carry a pager that would beep occasionally to indicate when they are to report their feelings and opinions about the ropes course that they are participating in. Or, instructors might randomly stop normal activities throughout the program and ask all of the participants to provide information about their experiences. Written questionnaires, journals or tape recorders can be used to gather information at randomly selected times. (See the sections on each of these techniques for a further explanation.)

Advantages

+Flexible time frame. This may be a good choice for evaluations which are conducted by the program's staff because all of the preparatory work could be done ahead of time, and the analysis could be postponed until after the session is over. Therefore, the evaluation staff could concentrate on other duties during the session.

+Unique and innovative.

Disadvantages

-Interference with the program.

-Demanding for the participants.

-Relies on the respondent's honesty and willingness to participate.

-Data analysis may be complicated if the directions have not been clearly specified.

More Information


Experience sampling is an innovative method for gathering data from program participants at random times during their experience.
Technique: Outside Observer

"Outsiders," who visit the program can be invaluable in providing an objective view of your program. This concept was made clear to me one day as a friend pointed out interesting trees, beautiful flowers, and unique rock formations along one of my usual hiking trails. Why had I never noticed these things? Later, I realized that my familiarity with the site had caused me to stop paying attention to the details of my surroundings.

Selective perception can happen when we observe our own program, too. We may have difficulty distancing ourselves from our perspective, which may prevent us from gaining a complete understanding of the program and its effects. A knowledgeable outsider(s) will often see both positive and negative aspects that even an attentive staff evaluator would miss. An outsider can often provide significant information on unintended effects. A team of outside observers can produce even more information.

This technique differs from hiring a professional evaluator to conduct an evaluation or to provide an expert opinion. Visiting observers are simply people knowledgeable about outdoor education who visit the site for a short period and prepare a report on their observation. They may use tools such as check sheets or tape recordings to aid in recording their observations. The primary information reported concerns the process, as opposed to the program's input or output. Whom should you ask to be an observer? Consider: educators, administrators of other programs, school principals, and university professors of outdoor education. The primary criteria for these observers are that they are not stakeholders, are genuinely interested in helping to improve the program, and will strive to present an unbiased report.

Advantages
+ Requires no time or expertise of program staff.
+ May be inexpensive (dependent upon the stipend awarded to the visitor and any equipment that is needed).
+ Offers an additional perspective, which reduces the threat of selective observation.
+ May provide information on unintended effects.

Disadvantages
- May be incomplete. By itself the use of outside observers probably would not constitute a complete evaluation, but it works well when combined with other methods.
- Relies on the objectivity of the outsider(s).

Visiting outside observers can make a substantial addition to your evaluation by making valuable suggestions, giving constructive criticisms, and providing data for triangulation.
Technique: Q-Sorting

Q-sorting is a method used to gather information on how people prioritize items such as program purposes or needs of the program. For example, the director of a retreat center might want to gather information about the needs of potential clientele for certain physical facilities. Through informal discussions with potential clientele, the director could identify some of the possible needs and could draft a list of the needs. After statements concerning each of these needs have been written on index cards, potential clientele could be asked to arrange the cards in order of importance. By providing blank cards and asking respondents to add any other needs that they might have, the director could discover needs not previously considered.

This technique can be used to gather information for a variety of attributes such as the importance, preference, or quality of program facilities, activities, or goals. To administer this procedure you must establish the evaluation questions, identify the appropriate respondents, and develop the statements. Since q-sorting gathers information about the stakeholders’ priorities, it is well suited for examining the program’s purposes. Q-sorting may be used within a group setting and may be appropriate for use with a Delphi technique which develops a group consensus. (See the information on Dialectic models, p. 40.)

Advantages

- Flexible. This seemingly simple technique can produce a variety of information. For example: (1) the overall rankings of each statement; (2) the consistency of the rankings among all respondents; (3) the degree to which a certain group of people ranks items similarly (i.e., do all of the parents rank nutrition as a great need); (4) the difference in the rankings of different groups (i.e., do staff and visiting teachers have opposite views on the importance of developing outdoor living skills); or, (5) the agreement of an individual’s ranking with an ideal distribution (i.e., for a religious program, do participants rate “living with Christ as my example” as a more significant program purpose then “living to please myself”).

- Pre/posttest possibilities. This could yield information on changes in attitudes and values.

- Straightforward analysis. You can use computers to compile frequencies and means of the quantitative results.

Disadvantages

- Analysis of participant-written statements. If the respondents create their own statements, you will need to develop a systematic method for the interpretation of the personal statements.
MISCELLANEOUS TECHNIQUES

More Information

Example:


\[G\text{-}\text{sorting offers a method for determining priorities or ranking values. It is relatively inexpensive and easy to administer.}\]
Technique: Self-Evaluation

Self-evaluation reports may provide valuable data for program evaluations. Self-evaluations can be recorded verbally or in writing. Often self-evaluation is used (1) to help an individual assess progress toward personal goals, (2) to determine how the perceived behaviors or attitudes of individuals have changed, or (3) to promote self-reflection. Personal data can be pooled together and analyzed for information concerning the program’s effects on a group. Findings from this type of analysis often include changes in self-esteem, locus of control, social maturity, attitudes about the program, and unintended effects of the program.

Advantages

+ A learning experience in itself. If the evaluator stresses that this is not a test, is a good listener, and demonstrates interest in the person as a developing individual, then self-evaluations in both the written and verbal form can be an exciting learning experience for both parties.
+ Qualitative data.
+ In-depth analysis of the personal growth and changes in staff and participants.

Disadvantages

- Requires trust. Initially, people are self-conscious and are not willing to freely discuss their strengths, weaknesses, and progress toward a goal. If a trusting relationship is not built, the data will be shallow and biased toward what participants think they should be experiencing and accomplishing.
- Necessitates good relationships. This technique relies heavily on the abilities of the evaluator and the receptivity of the participants.
- Hawthorne effects. Being particularly susceptible to these effects, this method should not be used if the results must be unquestioned.
- Highly demanding on the participant. It may be considered an interruption or a distraction. (However, the personal benefits often outweigh this concern.)
- Limited to participants who can be honest and analytical about their own actions and feelings.
- Time-consuming (and considerably more so if data are collected verbally in individual interviews).

More Information (See p. 101, the artifacts section for more on journal analysis.)
Technique: Sociometric Method

Sociometric methods are useful for examining the social structure of a group. This technique asks the respondents to identify their top choices of other students to be their 'partners' for some activity. These choices are then plotted on a sociogram which allows the evaluator to discern the group dynamics. For example, in the sociogram below, it becomes clear that Juan is a top choice for all of the children and that the boys, Juan, Jose, and Tommy would choose each other exclusively. This provides a visible picture of the social preferences within a group and makes for easy detection of isolates, cliques, and patterns of integration. Sociograms can be used several times throughout a long session, or before and after a program to evaluate the program's effect on group cohesion or personal relationships within the group. (The second option may be only feasible when the group is intact before and after the visit, as with a school group.) Sociometric methods are best if used in conjunction with actual observations of behaviors.

**Advantages**
- Is easy to administer and to analyze.
- Requires little time, money, or expertise.
- Can be used with young children.
- Provides input, process or output data depending on the timing of the collection.

**Disadvantages**
- Is limited to collection of information on social interactions and measures only one choice at a time.
- Requires input from participants.

**More Information**
Technique: Standardized Indicator

Standardized indicators have been developed and used by evaluators and researchers for years to "test" people for specific attributes. Indicators are usually in the form of written instruments developed to quantify attitudes, skill levels, or knowledge. An example of a standardized indicator is the "Children's Version of the Family Environment Scale" which, through thirty pictorial multiple-choice questions, assesses the child's perception of their family. They can be used in program evaluation to determine the outcomes of the program, and may be used to make comparisons of either a single group pre/post visit or between a group which has and one which has not visited the center.

Advantages

+ Produces "hard" evidence.
+ Yields standardized results. You can easily compare the results to other programs that use the same indicators.
+ Is trusted by the general public.
+ Is already constructed and tested for validity and reliability.
+ Is easy to administer and to analyze (which can often be done by a professional testing service for a fee).
+ Requires only a minor time commitment.

Disadvantages

- Requires certain assumptions:
  (1) The phenomena are numerically measurable (i.e., attitudes and feelings are quantifiable). Some outdoor educators are uneasy with this technique because it does not account for the context in which the attitudes are held and it is based solely on quantitative data.
  (2) Everyone will read and interpret the directions and questions in the same way. Reading ability, dislike for standardized forms, etc. may affect the validity of the results.
- Demands participants' time and energy.
- Is generally viewed as unpleasant.
- Causes an interruption. Typically the standardized, written, and non-personal format is counter to the practices of outdoor education.
- May be expensive.
- Is designed for a specific type of respondent (i.e., standardized indicators which are designed for adults may not be appropriate for children).
Figure 14 lists of some of the standardized indicators that have proven useful in assessing outdoor education. Addresses for the authors or publishers of some of these indicators are included in Appendix 8.

**FIGURE 14. Standardized Indicators**

[Note: Prices and other details of these instruments are included only as general references and are subject to change.]

<table>
<thead>
<tr>
<th>TOPIC</th>
<th>RESPONDENT</th>
<th>INSTRUMENT</th>
<th>REFERENCE</th>
<th>NOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>Participant</td>
<td>State-Trait Anxiety Inventory</td>
<td>Spielberger, Gorsuch, &amp; Lushene (1970)</td>
<td>Grades 9- adult, Two 20 item scales.</td>
</tr>
<tr>
<td>Attitudes towards people experiencing handicaps</td>
<td>Staff</td>
<td>Lazar Attitude Toward Handicapped Individuals Scale</td>
<td>Lazar (1973)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parents, Social workers</td>
<td>Comprehensive Assessment of Mastery and Problems</td>
<td>Deysach, Ross, &amp; Hiers (1979)</td>
<td>Four main skill areas: self-help, social, relationship with authority, and program of group skills.</td>
</tr>
<tr>
<td></td>
<td>Parents, Social workers, Teachers, Staff</td>
<td>Self-Control Rating Scale</td>
<td>Kendall &amp; Wilcox. (1979)</td>
<td>33 items.</td>
</tr>
</tbody>
</table>
### FIGURE 14. (Continued) Standardized Indicators

<table>
<thead>
<tr>
<th>TOPIC</th>
<th>RESPONDENT</th>
<th>INSTRUMENT</th>
<th>REFERENCE</th>
<th>NOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environmental Attitude</td>
<td>Participant</td>
<td>Environmental Knowledge and Attitude Inventory</td>
<td>Weigel &amp; Weigel (1978)</td>
<td>For high school students.</td>
</tr>
<tr>
<td></td>
<td>Participant</td>
<td>Millward-Ginter Outdoor Attitude Inventory</td>
<td>Millward (1973)</td>
<td>The specifics of the content limit use to northeastern U.S.A.</td>
</tr>
<tr>
<td></td>
<td>Participant</td>
<td>Weigel &amp; Weigel Environmental Concern Scale</td>
<td>Spacht (1980)</td>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
<td>Participant</td>
<td>Billings Environmental Assessment Instrument</td>
<td>Oloke (1981)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Participant</td>
<td>Environmental Knowledge and Attitude Inventory</td>
<td>Weigel &amp; Weigel (1978)</td>
<td>For high school students.</td>
</tr>
<tr>
<td></td>
<td>Participant, Parents, &amp; Staff</td>
<td>Problem Solving Decision-Making Style Inventory</td>
<td>Hersey &amp; Natemeyer (1982)</td>
<td>Self inventory &amp; inventory completed by others. Comparative scores, good for group use. Forms $2.95 each.</td>
</tr>
<tr>
<td>Leadership</td>
<td>Participant</td>
<td>Outdoor Adventure Education Situation Reaction Model</td>
<td>Badger (1979)</td>
<td></td>
</tr>
<tr>
<td>Locus of control</td>
<td>Participant</td>
<td>Novicki-Strickland Locus of Control Scale</td>
<td>Novicki &amp; Strickland (1973)</td>
<td>40 items. Yes/No response categories.</td>
</tr>
<tr>
<td></td>
<td>Participant</td>
<td>Rotter's Internality-Externality Scale</td>
<td>Rotter (1966)</td>
<td>29 items.</td>
</tr>
<tr>
<td>Topic</td>
<td>Respondent</td>
<td>Instrument</td>
<td>Reference</td>
<td>Notes</td>
</tr>
<tr>
<td>---------------</td>
<td>------------</td>
<td>-------------------------------------------------</td>
<td>--------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Personality</td>
<td>Participant</td>
<td>Children's Personality Questionnaire</td>
<td>Porter &amp; Cattell</td>
<td>Ages 8-12. Untimed, 30-60 minutes. Handkey, Scoring and Interpretation services available. Available in Spanish and German.</td>
</tr>
<tr>
<td></td>
<td>Participant</td>
<td>Sixteen Personality Factors Questionnaire</td>
<td>Cattell, Eber, &amp; Tatsuoka</td>
<td>Grades 3-up. 5 different forms. Available in Spanish, German, and on audio tape.</td>
</tr>
<tr>
<td>Problem solving</td>
<td>Participant</td>
<td>Outdoor Adventure Education Situation Reaction Model</td>
<td>Badger</td>
<td>Self inventory &amp; inventory completed by others. Comparative scores, good for group use. Forms $2.95 each.</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>Participant</td>
<td>Adjective Checklist</td>
<td>Gough &amp; Heilbrun</td>
<td>Ages 8-16 or 16-adult. 58 items. Untimed, 15 minutes. Manual=$6.00; 25 booklets for young=$5.00, Key=$2.00; 25 booklets for adults=$13.50, Key=$1.25.</td>
</tr>
</tbody>
</table>
### MISCELLANEOUS TECHNIQUES

**FIGURE 14. (Continued) Standardized Indicators**

<table>
<thead>
<tr>
<th>TOPIC</th>
<th>RESPONDENT</th>
<th>INSTRUMENT</th>
<th>REFERENCE</th>
<th>NOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participant</td>
<td>Feelings of Inadequacy Scale</td>
<td>Janis &amp; Field (1959)</td>
<td>23 items. Likert-type scale.</td>
</tr>
<tr>
<td></td>
<td>Participant</td>
<td>Index of Adjustment &amp; Values</td>
<td>Stremba (1977)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Participant</td>
<td>Lipsett Self Concept Scale</td>
<td>Lundegren (1975)</td>
<td>Grades 4-6. Appropriate for special populations. Five point Likert-type scale.</td>
</tr>
<tr>
<td></td>
<td>Participant</td>
<td>Perceived Freedom in Leisure Short Form</td>
<td>Ellis &amp; Witt (1986)</td>
<td></td>
</tr>
</tbody>
</table>
### FIGURE 14. (Continued) Standardized Indicators

<table>
<thead>
<tr>
<th>TOPIC</th>
<th>RESPONDENT</th>
<th>INSTRUMENT</th>
<th>REFERENCE</th>
<th>NOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social Climate</strong></td>
<td>Participant</td>
<td>Cultural Attitude Scales</td>
<td>D'Agostino (1990)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Participant</td>
<td>Outdoor Adventure Education Situation Reaction Model</td>
<td>Badger (1979)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student</td>
<td>Teacher Pupil Relationships Inventory</td>
<td>Knoblock &amp; Goldstein (1971)</td>
<td>For elementary school students.</td>
</tr>
</tbody>
</table>

**More Information**

Technique: Test

For years, tests have been the traditional method for assessing educational programs, and thus they have been well accepted as valid and reliable sources of information about a person's knowledge, mastery, or understanding. [Note: we use the term "tests" to refer to tests which are developed in-house. Standardized tests are discussed in the Standardized Indicators section.] Tests come in many forms, including: essay, true/false, multiple choice, matching, completion, or problems. Nevertheless, their use has been contested as limiting, as ignoring the effects of the context (the student's environment) on the results, and as favoring other skills such as test-taking ability and reading skills over the skill or knowledge of interest to the evaluator.

Advantages
- Produces "hard" evidence.
- Allows for assessment of individuals, groups, individual elements of the program, or the entire program.
- Is inexpensive (as a result of the in-house design).
- Is designed to assess specific program objectives.
- May incorporate one or many subjects, in which case the success of individual lessons or of the integration of the lessons can be assessed.
- Has the potential for strong validity. The validity of the results can be strengthened by providing additional supporting data collected through other techniques that can capture actual applications of the knowledge.

Disadvantages
- May be inappropriate for measurement of some attributes.
- Provides only information about the program outcomes. While this works for summative evaluations, tests usually do not provide enough information about the program processes to lead to program improvement.
- Is demanding and unpleasant for the respondent.
- Interrupts the program.
- Lacks incentives. When tests are conducted solely for program evaluation purposes, students may lack motivation to answer correctly.
- Requires comparisons. In order to document increased knowledge of the participants, comparison of either pre/posttest scores or of experiment and control groups is necessary. This requirement demands increased effort and may indicate the need for two equivalent forms of the test if pre/post comparisons are to be made.
- Does not account for all learning styles.
- May be culturally biased.

More Information

Tests can provide "hard" data, are relatively inexpensive, and can assess outcomes.
Appendices
Appendix 1: Validity and Reliability

Evaluators hope that the information which they collect is completely accurate. However, problems with either the manner in which the data are collected or with the information that people supply will cause some degree of error in the results. This is simply one more element of evaluation. In this appendix some approaches for reducing error are discussed.

UNDERSTANDING ERROR

In order to minimize error, we must first understand it. There are two basic types of error: errors in the validity of the data and errors in the reliability of the data. Validity refers to how correct or accurate the results are. For example, a conclusion is invalid if it states that a program provides accessible facilities when actually it has no accommodations for persons who use wheel chairs. Reliability refers to how consistent the results are. An example of unreliability would be if the same test were given to the same student twice under the same conditions (within a reasonable time frame), and the students' scores were very different.

Imagine that you are shooting arrows at a target. If you shoot with accurate validity, then all of the shots will hit the bull's eye; if your shot is not accurate, the shots will be spread around the target. If you have a reliable but not a very valid shot, all of the shots will be clustered together (i.e., getting the same results each time), but none of them will hit the bull's eye. You need both good validity and good reliability to support a successful evaluation. You can do this by ensuring the validity and reliability of both the evaluation design and of the instruments that you use. [Note: The term 'instruments' means anything that you use to collect data, including questionnaires, tests, human observers, etc.] We will first look at increasing validity, and then at threats to reliability.

FIGURE 15. Validity and Reliability

Validity

There are two types of validity: internal and external. *Internal validity*, also called credibility, is a measure of how accurate the data are. *External validity* refers to whether or not the results can be generalized to other situations.

The internal validity is threatened if there is a fault in the instrument or if the participants are not providing accurate data. If the instrument is faulty, it would be similar to reading a compass without compensating for magnetic versus true north. You would always be lead off your course by the incorrect reading.

In the case of an evaluation instrument, problems may result from the content, construct or predictive validity of the instrument. *Content validity* deals with whether the instrument measures all of what it is supposed to and *only* what it is supposed to. For example, if you want to test a student’s problem-solving skills, your instrument should not include items (questions) which require rote memory or high level reading ability. When a test does require other skills, the scores may be more a reflection of the other skills than of their problem-solving abilities. An instrument which has good *construct validity* addresses all of the individual parts of the topic that are being measured. In other words, if you want to determine a person’s attitude toward the environment, and it is accepted that this attitude is made up of certain elements (i.e., feelings towards animals, concern for habitats, opinions about pollution), then a valid instrument will measure each of these elements. *Predictive validity* refers to the ability to use the results from one instrument to predict the score on another instrument which measures the same thing. If your study concludes that campers’ self-esteem improved as a result of the program, yet accepted tests of self-esteem and the observations of parents conclude that it decreased, there may be a problem with the predictive validity of your instrument.

If the data that you collect are incorrect or incomplete, the credibility of your study will be questionable. In some cases, respondents may intentionally provide false information. They may supply only socially desirable answers (i.e., I never litter.). Or, they may just not respond if they have negative feelings toward the program. If respondents behave differently because they know that they are being observed, then the results will not be valid. This reaction is called the *Hawthorne effect*. This is named after a research case where the productivity level of factory workers was found to be more dependent upon the workers’ knowledge that they were part of a study, than upon the working conditions in the factory. Hawthorne effects can cause serious problems with the internal validity of an evaluation.

As previously mentioned, if a pretest is given, it may contaminate the results of the posttest. In these cases, the final scores may represent only what the students know about the particular questions or topics which are on the test, but the test is not a valid measure of all of the material that was learned or experienced in the program.
Another concern about internal invalidity relates to the conditions that surround the participant when the evaluation is conducted. A person's environment may influence the way that the individual responds to questions or how the person acts while being observed. For example, imagine the results of an interview with a staff member in which the individual is intimidated by the presence of upper level administrators. Or, consider how willing parents would be to answer a questionnaire when they are about to pick up their child from his/her first camp experience? It is crucial to consider how a person's environment might affect the evaluation results.

You can eliminate many of these threats to internal validity as you plan your evaluation. Test your instrument by having some of the stakeholders (who will not be a part of the actual study) read and comment on its content validity. Ask a panel of experts (staff or other directors) to rate its content and construct validity. [Note: Be sure to provide clear instructions to the panelists.] To measure the predictive validity of a test, find another test that is on the same topic. Have one group of people complete both tests and compare the scores.

Ideally, all pre/posttests should be conducted under identical conditions. To minimize contamination of the posttest by the pretest, use a different form of the test as a pretest, or administer the pretest well in advance of the program. To reduce error, make use of an anonymous or hidden observer, use video tapes, or make lengthy observations (participants rarely keep up their "best behavior" for extended periods of time).

If you use a person as your instrument, take special care concerning the validity. The best way to reduce error resulting from observer bias is for evaluators to state their bias and to try to determine what effects the bias might have on the results. Repeating and recording observations (with a tape or a camera), plus allowing participants to cross-examine the observer's conclusions, will bolster the observer's credibility.

The internal validity of every evaluation can be improved by using multiple instruments. This is termed triangulation, and it allows you to collect information from several different perspectives. One might interview staff, observe participants, and read program records, to gain a better understanding of the safety procedures practiced in your program. Although this requires additional work, it has great potential for catching data that otherwise might have been missed and for improving the accuracy of the conclusions.

The external validity of a study is important for the application of the results. If you want to apply the results of this summer's evaluation to next summer's program, you will want to know that your conclusions can logically be transferred to another situation with similar conditions. One of the greatest causes of external invalidity is the lack of a representative sample. Most evaluations gather information from only a limited number of stakeholders (a sample representing the entire group), since contacting all of the stakeholders is not feasible. If the sample adequately represents the entire group, then you can assume that the information that you gather from the sample may the same as that which you would have gotten from the larger group. (More information on selection of a representative sample can be found in
Appendix 3] [Note: Since the conditions from season to season may vary tremendously, it may be very difficult to generalize from one season to the next.]

Ensuring that your sample is representative is one step to improving external validity; however, this does not guarantee that the results will be applicable to other programs. Many naturalistic researchers believe that since no two social situations can ever be exactly alike, total transferability of results from one program to another is impossible. Instead of making sweeping statements about the generalizability of their findings, these evaluators generally prefer to write reports which include “thick” descriptions of contextual detail. This allows the reader of the evaluation report to compare the situations and determine what information might be useful in their unique situation. (See Henderson, 1991.)

Reliability

The reliability of an instrument is a measure of its consistency. If you use a numerically scored instrument, such as a standardized test, then you can determine a reliability coefficient (expressed as a decimal between 0.0 and 1.0) for your instrument. This allows you to compare your instrument’s reliability with that of other similar instruments. For example, if you give a group of students a reliable test one day and then again on the next day, all should receive approximately the same scores, and the reliability coefficient would be a 1.0. If the students received drastically different scores, then the reliability coefficient would be approximately 0.0. By testing and retesting, a small sample of people (20-40) who will not be a part of the evaluation, you can determine the reliability of your instrument. Computerized statistical programs such as SPSS facilitate these calculations. Items which are not reliable should then be modified or removed from the instrument.

Instead of the test-retest method, naturalistic evaluators use audit trails to support their reliability. An audit trail should include a thorough description of the procedures, the decisions made as the evaluation unfolded and an explanation of the logic that was used to draw conclusions. The audit trail should illustrate how another evaluator, faced with the same data, would conduct the evaluation in the same manner. In addition, a second opinion can be used in the interpretation of the data in order to ensure reliability and validity. (See Henderson, 1991, for more details.)

Both sound validity and reliability are crucial to eliminating error in any evaluation. A consistent instrument may not be an accurate one. Always check your instruments and design for all types of threats.

To reduce error in your evaluation, remember to:

(1) clearly document the evaluation process,
(2) consider the effects of the situation on the subjects,
(3) establish the validity and reliability of your instruments, and
(4) acknowledge your biases.
Appendix 2: Data Collection Times for Quasi-Experimental Designs


Quasi-experimental-type designs require data collection at specific times. A number of specific designs are described below which differ in the timing of the data collection. Advantages and disadvantages of each are listed. [Note: Although formal tests are often utilized with this design, the words pretest and posttest are used in this text refer to any form of data collection.]

**Single Case Study**

Program..........................Posttest

Single case studies include only one data collection during or at the end of the program or event. Most traditional camp surveys and written tests follow this type of design. No comparisons are made. The time commitment is intense but short.

**ADVANTAGES:**
+ Low cost.
+ Easy implementation.

**DISADVANTAGES:**
- Without a pretest, there is no proof of a change or that the program itself was the cause of any change.

**Immediate Pretest and Posttest**

Pretest............Program.........Posttest

Pretest data are collected prior to or as the participants arrive at the program. Posttest data are collected again during or after the program or event. A comparison is made between the pretest and the posttest data from the same group of participants in order to investigate the effects of the program or event.

**ADVANTAGES:**
+ Documents a change. You can look at the progress of each participant and/or the trends of the entire group.
DATA COLLECTION TIMES

DISADVANTAGES:
- Requires two observations (demanding on both the staff and the participants).
- Influenced by outside variables. Variables such as the excitement caused by arrival or departure, may lead to inaccurate information. This error must be carefully considered, especially when an attitude toward the program itself is being measured. The use of a delayed pretest and posttest can minimize these effects. Immediate collection of the data may be valid if an appropriate time is chosen.
- Indicates only that a change has occurred. This does not provide information about exactly why the change occurred or that this program causes more change than another program.
- Contaminated by pretest. Particularly if you are assessing knowledge gain, the pretest itself may affect the posttest scores. Pretests may alert students to topics of particular importance. They may then (intentionally or unintentionally) seek out that specific information and thus score higher on the posttest. Using alternate forms of the test for the two collections can help you to avoid this risk. Also, giving the pretest several weeks in advance may minimize what the students remember by the time of the program. If you want to determine how much effect the pretest is having on the posttest results, start with equivalent groups of campers. (Groups are considered equivalent if they have similar characteristics, e.g., comparable intelligence, economic backgrounds, previous knowledge.) Give one group both tests and the other group only the posttest. The difference in the groups’ posttest scores may be the result of the pretest.

Delayed Pretest and Posttest

This design includes two data collections from several days to several weeks before or after the participants come to the program. The important consideration is that there is an extended time commitment.

ADVANTAGES:
- Documents a change.
- Provides information on short-term retention of knowledge, attitude change, etc.
- Minimizes error caused by pretests.
- Offers a comparison of individual or group results.

DISADVANTAGES:
- Demands an extended time commitment.
- Requires access to participants prior to and after the program. (This may be appropriate for intact groups such as school classes.)
- Contaminated by pretest.
Follow-up Study

The emphasis of a follow-up study is in tracking long-term effects of a program through a single collection or a series of posttest data collections over an extended time period after the program.

**ADVANTAGES:**
- Measurement of retention and long-term effects. Comparisons can be made of the results after different periods of time.
- Can be used with immediate posttests to determine the rate at which the effects of the program are lost.

**DISADVANTAGES:**
- Extended time period and a long-term commitment are required.
- Participants may become difficult to contact or may lose interest in responding.
- Other events in the participants' lives may overshadow or interfere with the long-term effects of the camp program.

**More Information**
Appendix 3: Sampling

Ideally, evaluations should be census studies, i.e., collecting data from all individuals within the population (the group of individuals of interest). However, resources are limited and populations (i.e., all of the program participants in a year) are often large and dispersed, making it infeasible to perform a census study. Fortunately, the use of samples provides an efficient way to gather data from small groups of individuals in order to estimate the results that would be obtained through a census. There are two categories of samples: probability and non-probability. A probability sample is one in which all members of the population have an equal chance of being chosen. Thus, results from a probability sample can be generalized to all of the individuals in the population. A non-probability sample is comprised of individuals from the population who are not necessarily representative of that population. All sampling is susceptible to error; however, a properly conducted probability sample often provides adequate information and a well-designed, carefully interpreted non-probability study can provide close approximations.

SAMPLE SELECTION

Probability Samples

There are three basic types of probability samples, all of which include a random sampling process: (1) simple random sample, (2) systematic random sample, and (3) stratified random sample. All random sampling must begin with a list, called the frame, of all of the members of the population. This list must be complete, must not include any duplicates, and must not have any inherent structure that would bias selection. To create a simple random sample, number all of the individuals and randomly choose numbers (through a random number generator or by picking numbers out of a hat) until you have as many individuals as you need. To derive a systematic random sample, start by dividing the number of the total population by the number that you need in your sample. This will give you a number, r. Then, start at some random point in your frame, count off r numbers, and select that individual for the sample. Continue this process until you have all of the persons for your sample.

A stratified random sample is used when there are subgroups within your population (for example, people of different ages), and you want to insure that each subgroup is adequately represented. To do this you must determine how much of the population each subgroup, or stratum, comprises, and your frame must be broken down into these strata. Then create a simple random sample from each stratum such that the proportion of individuals in the sample from each sub-population is the same as it is in the population. From any of the probability samples you can estimate the sampling error, identify the likelihood of any individual being selected, and make sound generalizations to the population based on the results from the sample.

Non-probability samples

A non-probability sample often is chosen because it is convenient, i.e., the school group that happens to be at the center when the staff member has the time to make observations. This type of sampling, called accidental or chunk sampling, is the easiest but should be avoided because it lacks all control over outside influences (you have no idea if it is a group of gifted students or not) and thus is associated with a high risk of error.
Sampling

When a certain group of individuals is chosen for a sample because of unique characteristics (i.e., all of the participants who have never camped out overnight), that is a purposive sample. These are useful when the program's effects on a certain type of participant is of particular interest. In such cases, clear specifications of the subject selection process should be included in the report and limitations need to be placed on the generalization of the results. The third type of non-probability sample, a quota sample, is used when representative proportions of individuals are chosen from separate sub-groups within the population, but within the strata, individuals are not selected randomly. Whenever a non-probability sample is used, you must be careful not to make direct inferences to the larger population; you will not know the amount of sampling error, and you cannot determine the probability of selecting any individual to be in the sample.

Sample Size

In conjunction with choosing your sample, you need to determine the size of your sample. The size of the sample that you chose will probably depend on your resources and how accessible the population is. If you are conducting a controlled, quasi-experimental study, there are specific equations which you can use to determine the appropriate sample size. (See Kraemer & Thieman, 1987.) There are some generalizations which can help in determining sample size. First, if all of the respondents are similar to one another, there is a small likelihood of extreme scores, and a small sample size can be used. The more precision needed in the results, the larger the sample must be; as the sample size increases, the mean of the sample scores will come closer to that of the population. With stratified and quota samples you may need a larger total sample size in order to get adequate samples from each strata. And, if you are unwilling to risk the chance of getting an unrepresentative sample, then you must have a large sample. The last consideration is that of evaluation resources. An ideal sample is large, but you are on a tight budget, remember that it is better to concentrate your resources on a thorough study of a small number of people than it is to spread your resources too thin to conduct a broader evaluation.

More Information

Appendix 4: Non-Response Error

Non-response error occurs when data (that are collected from a sample of the population) are not representative of the entire population because individuals with unique characteristics within the sample chose not to provide responses for the evaluation. Compare this to looking for ripe strawberries in a field. If you go to the edge of the field and pick five berries, taste them, realize that they are not yet ripe, and then leave the field hungry, you may be leaving behind some delicious berries. What if the few berries that you picked, did not get as much sunlight as the rest, and therefore were not as ripe? Certainly a hungry person would not go away without sampling from several locations. Similarly, if you get responses from only half of your sample, you should not be satisfied that they are representative of the entire population. Perhaps only those who were pleased with the program decided to reply; conclusions drawn only from these respondents would obviously be biased. The survey techniques (mailed questionnaire, interviews, and focus groups) are particularly susceptible to having non-response error. Yet, when properly conducted, these methods can result in very high response rates and in some situations, there are ways to assure that the respondents are in fact representative of the population.

The first step to reducing non-response error is gaining a high response rate (at least 80%). This can be accomplished through designing your instrument (questionnaire, focus group questioning route, etc.) with the respondent in mind and through the inclusion of a follow-up procedure within your design. (See Appendix 6 for hints on successful instrumentation.) To deal with the non-respondents, you have three basic choices: (1) ignore the non-respondents, (2) legitimize the respondents as representative of the population, or (3) verify that the non-respondents have no unique characteristics, thus the respondents are representative of the population.

The first choice will result in weak validity of the data, but may be necessary where the need for exacting conclusions is low, there is no indication that the non-respondents differ from the respondents, or when resources will not allow further efforts. In such cases, the evaluation report should stress that the conclusions apply only to the respondents. If you compare the respondents to known characteristics of the population and can confirm that the respondents are indeed representative of the population, you can generalize from the sample to the population. To verify that the non-respondents are not unique, you can compare them to the respondents on known characteristics. Or, you can track down a random sample of the non-respondents (10%-20%), possibly though a telephone interview; determine the mean results for that sample; and, compare them to the original respondents' results. If these data are comparable, then you can assume that the non-respondents do not represent a unique sub-population. When you cannot substantiate through these measures that the respondents are representative, you must limit all conclusions as applying only to the respondents.

More Information

Appendix 5: Content Analysis

Data that you have collected can become intimidating if you are not familiar with data analysis. If you have a large quantity of tape recordings or written responses, you may want to use a content analysis to help you interpret the data. "Content analysis" is any of a number of systematic methods for analyzing a document that results in concise objective information. One way to analyze a document (a journal entry, a transcript of an interview, etc.), begins with the development of a coding system for the topics addressed. For example, all sentences about counselors might be coded with a 1 and all of these sentences which indicate negative opinions toward the counselors might be coded as \(1N\). The next step requires studying the document and coding each segment. Finally, you can categorize and tally the coded information. This type of analysis requires an intricate numerical coding system and could result in a conclusion that is supported by numerical data. If you choose this type of analysis, we highly recommend that you use a computer software program, such as The Ethnograph, to simplify the coding and sorting of data.

There are other more qualitative methods of analysis that might be useful. For example, you might study the data to detect patterns or recurring themes (i.e., in the focus groups, parents kept referring to the poor nutrition provided at the program). You would then base your conclusions on these themes and could support them with direct quotations from the respondents. All of the above variations on analysis have proven useful in different situations; your choice of approach should be based on the type of results that will be most useful and applicable to your program.

More Information

Example:

Appendix 6: Hints for Successful Evaluations

Successful evaluations require the use of successful instruments. In order to be successful, these instruments, whether they are questionnaires, interview schedules, or academic tests, must 1) appeal to the respondent, 2) provide valid and reliable data, and 3) produce analyzable data. Careful consideration of these three characteristics prior to administration of your instrument can save you from wasting precious resources on an unsuccessful instrument.

The data's validity, reliability, and ability to be analyzed should be tested in pilot and field tests conducted prior to the actual administration on people who are similar to, but not a part of, the study's sample. Problems with face validity, vocabulary, and layout of the instrument can be detected. Your field test is a "trial run" of the actual data collection process and should be conducted exactly as if it were the actual collection. This will help identify problems of face validity and analysis of the data as well. Additionally, if you ask the same individuals to complete another instrument after a short period of time, then you can compare the difference between each individual's set of responses to determine if your instrument gives the same results each time. This test-retest procedure allows you to determine the instrument's reliability.

Before you can do any of this testing on the validity, reliability, etc. of your instrument, however, you need to make sure that you have developed a user-friendly instrument. When designing the instrument, place yourself in the respondent's shoes. You need to understand your population well enough to know what appeals to and what offends them. Consider their concern for the time and effort required, their interest in the evaluation, and their feelings about the confidentiality of their responses.

First, you must arouse the respondent's interest in the evaluation. All of the respondents will have some personal interest in the program; it is your task to make them realize that it is personally advantageous to them to help improve the program. This can be accomplished by 1) reminding them of their experience (through pictures, stories, etc.), 2) demonstrating that their responses are valued highly, 3) reporting how many others they were chosen to represent, and 4) stating exactly how their involvement will improve the program for their future use. You may want to give specific incentives for participation in the evaluation. Some examples are: money (a small monetary gift so that they can enjoy a beverage while they complete the questionnaire), candy, pencils, a ball cap, a tee-shirt, stationery, or a gift certificate from your camp store.
HINTS FOR SUCCESSFUL EVALUATIONS

Second, recognize the time and effort that each respondent is being asked to give. Ask yourself, "If I were asked to be a respondent, would I? What would encourage me? What could discourage me?". To decrease the demands on the respondent, there are a few things that you can do:

(1) state directions and expectations clearly,
(2) design all written material so that they look easy and are easy to read - include lots of white space in the overall layout to avoid cluttering and confusion:
   • use large print;
   • provide stimulating pictures, colors, or other graphics;
   • use spacing, boldface, and underlining to separate questions and to emphasize directions;
   • have all questions flowing down the page and response categories to the right or under the questions;
(3) use unbiased and concise language (vague words such as "few" and "young" have many interpretations, so try to avoid these or include corresponding numbers);
(4) if you use numerical response categories (as in a rating scale), insure that their meaning and the direction of the rankings are understood;
(5) make sure that the questions and vocabulary are appropriate to the respondent (it is frustrating to be asked questions that you cannot answer);
(6) be courteous, respectful, honest, and appreciative;
(7) always provide space for additional comments or responses that are not listed.

In addition to these practical hints for the actual instrument, consider the following recommendations for your design:

(1) use pre/post-contacts (postcards or telephone calls) to establish your credibility, to build a positive relationship, and to inform the respondent of the purposes or results of the evaluation;
(2) continually check your design against the needs of the evaluation audience:
   • are the important needs being met?
   • are the plans feasible?
   • will the results be useful?
(3) ask yourself if your techniques are appropriate for your population;
(4) acknowledge your biases and assess their effects on the evaluation;
(5) determine and report the degree of validity, reliability, and bias of the instruments;
(6) always be flexible and creative (the best information may come from sources that were not even considered originally);
(7) in every situation, look for unintended effects and evidence that is counter to your expectations (this will improve the validity of your results);
(8) be happy that you are on your way to substantial program improvement!
Appendix 7: Evaluation Criteria for School Related Programs


PHILOSOPHY

1. The activities should maximize the use of the natural, human-made and human resources available at the site (i.e., activities which can not be taught as well at school).

2. The program should be conducted during the school year and not on vacations and holidays.

3. The students should become involved in the lessons through questioning, discussion, experimenting, evaluating, communicating, and problem solving whenever possible. Lectures by teachers and reading by students should be kept to a minimum.

4. To some extent, students should be able to choose some of their learning activities and some methods for meeting the specific objectives.

5. All activities should emphasize: 1) cooperation over competition, 2) successful experiences over frustrating experiences, 3) intrinsic rewards over extrinsic rewards, and 4) personal strengths over personal weaknesses.

PREPLANNING

6. Adequate time should be allowed for each activity.

7. Groups size should be small to maximize student participation and hands-on learning.

8. Both the teaching and the administrative staff should plan the course content and teaching methods which are designed to meet the lesson objectives.

CURRICULUM

9. The activities should be designed to meet the goals and objectives of the school curriculum.

10. All activities should be compatible and consistent with each other.

11. The activities should result in the preservation and/or the improvement of the natural and cultural environment.

12. A variety of educational objectives should be emphasized, including those which pertain to knowledge, skills, and values.

13. The benefits from the program should be worth the costs involved (e.g., time, energy, and money).

14. Interpretations of local history, ecology, etc. should be accurate and authentic.
EVALUATION CRITERIA

15. Student needs and interests should be considered, and the activities should provide opportunities for challenge and adventure whenever possible.

16. The curriculum of the program should be integrated with the curriculum of the classroom before and after the visit.

17. The activities should be appropriate for and recognize the individual nature among students.

LEADERSHIP

18. Teachers, administrators, and other school personnel should be given adequate in-service training through workshops, courses, and independent study.

19. The resident outdoor education program should use the best available personnel in the school system and the surrounding area.

20. Personnel should be qualified and trained to assure the best possible instruction for, treatment of, and safety for the students.

INSTRUCTIONAL MATERIALS AND FACILITIES

21. Adequate supplies and equipment should be available.

22. The facilities and services should serve the needs of the program without undue inconvenience or disturbance.

EVALUATION

23. The program should be constantly evaluated by the stakeholders and should be revised on the findings of the evaluations.
Appendix 8: Addresses for the Publishers and Authors of Selected Standardized Indicators and Computer Software

The following is a list of addresses for the publishers or authors of some of the standardized indicators referred to in the text and for the recommended computer software. This list is not complete, but it may aid you in locating a specific item of interest. Note that the costs and shipping times of the standardized indicators vary.

Adjective Checklist
Consulting Psychologists Press, Inc., 577 College Ave., P.O. Box 60070, Palo Alto, CA 94306.

Billings Environmental Attitude Assessment Instrument
Physical Science Department, Eastern Montana College, Billings, MO 59101.

Coopersmith Self-Esteem Inventory
Consulting Psychologists Press, Inc., 577 College Ave., P.O. Box 60070, Palo Alto, CA 94306.

Children’s Personality Questionnaire
Institute for Personality and Ability Testing, Test Services Division, P.O. Box 188, Champaign, IL 61824-0188.

Children’s Version of the Family Environment Scale
Consulting Psychologists Press, Inc., 577 College Ave., P.O. Box 60070, Palo Alto, CA 94306.

Culture-Free Self-Esteem Inventories for Children and Adults
Special Child Publications, P.O. Box 33548, Seattle, WA 98133.

Ethnograph
Family Environment Scale
Consulting Psychologists Press, Inc., 577 College Ave., P.O. Box 60070, Palo Alto, CA 94306.

Hahneimann Elementary School Behavior Rating Scale
Preventative Intervention Research Center, Philadelphia, PA 19102.

Piers-Harris Children’s Self-Concept Scale
Western Psychological Services, A Division of Manson Western Corporation, 1203 Wilshire Boulevard, Los Angeles, CA 90025.

Problem-Solving Decision-Making Style Inventory
Learning Resources Corporation, 8517 Production Ave., P.O. Box 26240, San Diego, CA 92126; (714) 578-5900.

Sixteen Personality Factor Questionnaire
Institute for Personality and Ability Testing, Test Services Division, P.O. Box 188, Champaign, IL 61820.
ADDRESSES

**SPSS**
Marketing Department, SPSS Inc., 444 North Michigan Avenue, Chicago, IL 60611; (312) 329-3500; fax (312) 329-3668.

**State-Trait Anxiety Inventory**
Consulting Psychologists Press, Inc., 577 College Ave., P.O. Box 60070, Palo Alto, CA 94306.

**Tennessee Self-Concept Scale**
Western Psychological Services, A Division of Manson Western Corporation, 1203 Wilshire Boulevard, Los Angeles, CA 90025.

**Walker Problem Behavior Identification Checklist**
Western Psychological Services, A Division of Manson Western Corporation, 1203 Wilshire Boulevard, Los Angeles, CA 90025.
GLOSSARY

**Actual Decision Makers**: those people who make and implement decisions about the program. People who work with the program daily (i.e., staff) may not be the authorized decision makers; however, they may have a strong influence on the daily events of the program.

**Affective Change**: a shift in an individual's attitudes, opinions, or beliefs.

**Anecdotal Recording**: an evaluation technique through which data are gathered from informal sources such as letters or personal conversations.

**Artifact Analysis**: an evaluation technique which relies on examination of items which participants have created, e.g., poetry, artwork, personal essays, etc.

**Assessment**: the process of estimating the value, worth or quality of some object or item.

**Audience**: those persons who will read the evaluation report and use the information provided.

**Audit Trail**: a detailed written record of the steps which were taken during an evaluation process.

**Authorized Decision Makers**: those people who, by virtue of their position, have the official power to make and enforce changes in program policy (for example, board members or administrators).

**Behavioral Change**: a modification in an individual's actions.

**Construct**: an idea which is composed of and refers to multiple simpler ideas or images, e.g., the term "camp" is a construct because it refers to many different types of programs.

**Construct Validity**: the extent to which an evaluation or an evaluative instrument assesses each component of the evaluation object.

**Content Validity**: "the representativeness of the instrument content to the topic being measured" (McCaslin & Others, 1979, p. 17).

**Decision Tree**: a written outline describing decisions to be made and possible solutions.

**Direct Observation**: an evaluation technique which requires watching and recording an individual's or a group's behaviors.

**Emergent**: without a formal plan; a design that evolves throughout the evaluation.

**Evaluation**: "a systematic investigation of the worth or merit of some object" (JCSEE, 1981, p.12).

**Evaluation Model**: an established approach or framework used as a guide to designing an evaluation.

**Evaluation Technique**: a method for collecting information that is to be used for evaluative purposes, e.g., personal interview, participant observation, or testing.

**Experience Sampling**: an evaluation technique in which, at random points in their experience, individuals are asked to provide information.
Experiential Education: the process of providing situations in which learning can occur through direct encounters with materials or learning by doing.

External Validity: a dimension of an instrument or evaluation's validity concerning the propriety of applying conclusions drawn from it to other situations.

Fixed: having a structured plan; an evaluation design that is detailed and adhered to throughout the evaluation.

Focus Group: an evaluation technique through which information is gathered from a group of individuals as they discuss a common topic.

Formative Evaluation: a systematic investigation of a program which is conducted while the program is in process, usually for the purposes of description and improvement.

Hawthorne Effect: a situation which causes error in data as a result of subjects knowledge that they are part of a study (e.g., participants are on their "best behavior" when an observer visits their cabin).

Importance: a term used in the Importance/Performance model to refer to how significant an item is to an individual.

Instrument: any tool which is used to collect information, including: written questionnaires, interview schedules, check sheets, human observers, etc.

Internal Validity: the accuracy of the information produced by an evaluation or an instrument. Content, construct and predictive validities are subcategories of the internal validity.

Interview Schedule: the list of questions and probing comments to be used in an interview.

Item: one element of an instrument, i.e., one question on a questionnaire, one type of behavior to be observed, one topic to be rated on a check sheet, etc.

KASA Change: change in an individual's knowledge, attitude, skills, or aspirations to behave in a certain manner (i.e., their intentions to change their behavior).

Meta-evaluation: an evaluation of an evaluation.

Moderator: a person who is trained to facilitate, observe, and record the discussion in a focus group interview.

Naturalistic: an approach to evaluation or research which seeks to obtain a holistic view of the uncontrolled program through inductive reasoning.

Non-probability Sampling: any process of selecting individuals to participate in an evaluation in which certain individuals have a higher chance of being selected.

Object: the person, thing or program element being evaluated.

Operationalize: to define a variable in specific, applicable terms that permits some type of measurement.

Outdoor Education: "education in, for and about the outdoors" (Smith, 1963, p.19).
**Performance**: a term that is used with the Importance/Performance model to refer to the level of achievement of the action being evaluated.

**Population**: the entire set of individuals from whom you would like to collect data (e.g., all of the participants or all of the leaders).

**Posttest**: a collection of data from participants after they have completed a program.

**Predictive Validity**: "the ability of an instrument to predict future performance" (McCaslin & Others, 1979, p. 17).

**Pretest**: a collection of data from participants prior to their visit to the program.

**Primary Data**: information that is gathered directly (e.g., direct observations of the activities included in an outdoor leadership session).

**Probability Sampling**: the process of selecting individuals to participate in an evaluation in which each individual has an equal chance of being selected.

**Probe**: a short comment which is used to encourage a more in-depth response from respondents, e.g., if the main question was "Why did you enjoy the over-night?" and the respondent replied, "It was fun," then a probe such as "What parts of the trip were fun?" might result in more in-depth information.

**Program**: an organized set of activities which have been designed to accomplish specific goals.

**Program Evaluation**: a description and an assessment of the occurrences within, the value of, and effects of an organized set of activities for the purposes of: improvement of or decision making about the set of activities.

**Purpose**: a statement about overall intentions to be fulfilled by a program. These are long-term and general in nature. They can be divided into more specific measurable objectives.

**Q-Sorting**: the process of rank ordering a set of written statements; used as an evaluation technique to gather information on priorities or opinions.

**Qualitative Data**: information that is collected in narrative form, typically descriptive of the quality of some item.

**Quantitative Data**: information which is collected in a numerical form.

**Questioning Route**: the outline of questions and probes used by the facilitator to structure a focus group interview.

**Reliability**: the degree to which an instrument or an entire evaluation process produces information which is reproducible if applied again under the same situations, i.e., consistency.

**Reliability Coefficient**: a numerical expression of the reliability of an instrument, expressed as a decimal between 0 and 1.

**Residential**: referring to a program at which the participants remain on the site for at least three days.
Residential Outdoor Education Programs: permanent programs where people are brought together in order to acquire knowledge; develop, modify, or clarify attitudes; learn skills; or, explore personal interactions through living together and participating in a planned course of events in an outdoor setting. These include, but are not limited to, private summer camps for children, year-around school camps, therapeutic programs, adult adventure expeditions, religious camps or retreats, and organization-sponsored camps.

R.O.P.E.: residential outdoor program evaluation; a process for designing an evaluation of a resident outdoor education program.

Sample: the subset of the population from whom you actually collect data.

Secondary Data: information that was originally collected for one purpose but is available for another secondary purpose. Compare to primary data.

Sociometric: evaluation techniques which gather information about social relationships.

Stakeholder: any person who has an interest in a program (i.e., participants, staff, founders, teachers, school boards, boards of directors, parents, the community, church or organizational leaders).

Standardized Indicator: a written instrument which has previously been developed and tested for reliability and validity.

Summative Evaluation: a final appraisal of a program usually used for decision-making purposes.

Test-retest method: the process of administering an instrument twice to the same individuals within a short time period in order to determine the reliability of the instrument.

Transcription: the process of making a written copy of data which has been recorded on tape.

Triangulation: the use of several evaluation techniques brought together to evaluate one topic.

Unobtrusive Observation: an evaluation technique which uses observation of individuals who are not aware that they are being observed for evaluation purposes.

Validity: the extent to which an evaluation or an instrument measures what it is intended to measure. Internal and external validity are two types of validity.
REFERENCES


REFERENCES


REFERENCES


"Interdisciplinary opportunities in a camp setting: Work/Service/Research/Education." Available from Mary Faeth Chenery, Department of Leisure Studies and Services, University of Oregon, Eugene, OR 97403.


REFERENCES


REFERENCES


Spacht, R.J. (1980). *The relationship between environmental concern and participation in a selected high adventure program*. (Doctoral dissertation, Indiana University). (University Microfilms No. 81-20,142)


Walsh, P.L. (1980). *An empirical, evaluative comparison of the goal-based and goal-free approaches to educational evaluation* (Published doctoral dissertation, University of Illinois at Urbana).
REFERENCES


