Criteria for Program Evaluation in State Motorcycle Safety Programs: State Administrator Perspectives

Raymond J. Ochs, Associate Professor
Traffic Safety Institute, College of Justice and Safety
Eastern Kentucky University

Abstract

This paper presents the results of a qualitative research study. The two primary research questions were: "What are the current program evaluation practices of state motorcycle safety administrators in their rider education and training program?" and "What would state motorcycle safety administrators recommend as ideal program evaluation practices for a rider education and training program?" Eight administrators from the southeast region of the United States were interviewed. The primary questions for the interviews were established using the Context-Input-Process-Product (CIPP) model of program evaluation. Two hundred and five criteria for program evaluation were identified from the interview process. In order to answer the two research questions, a double Delphi technique was employed, with a criterion selected if more than half of the administrators responded favorably as to a criterion's current use or its value in program evaluation. Seven of the eight program administrators responded to the subsequent surveys. A total of 17 criteria were identified in current program evaluation practice, and 30 criteria were identified for ideal program evaluation practice. Seven criteria were in both categories. They were 1) Capturing and acting on input from instructors, 2) Course participant evaluation results, 3) Degree of professional development among instructors, 4) Dollars spent on quality assurance, 5) Maintenance condition of the motorcycles, 6) Quality of instruction and instructors, and 7) Quality of instructor updates. A literature review provides a brief history of program evaluation, including a summary of motorcycle safety training and education research, and briefly describes several educational program evaluation models. It is recommended that exploration into ideal program evaluation criteria continue, that mechanisms for capturing information for the 30 criteria of ideal practice be explored, and that a model of program evaluation for motorcycle safety programs be developed.

Purpose

The purpose of this study was to identify criteria for program evaluation in motorcycle safety rider education and training programs from the perspective of program administrators. The study had two specific objectives: 1) to describe current program evaluation practices by state administrators, and 2) to develop program evaluation criteria based on administrators= recommended ideal practices. The primary research questions were: 1) What are the current program evaluation practices of state motorcycle safety administrators in their rider education and training programs? 2) What would state motorcycle safety administrators recommend as ideal program evaluation practices for rider education and training program?

Limitations of the study included the following: 1) The findings of this study are limited to the states where program administrators were interviewed, 2) The criteria for program evaluation are limited to the sample states, 3) The data collected was ex post facto relative to the CIPP program evaluation model, and 4) The evaluation criteria that resulted are from the perspective of the program administrators.

Research Methodology

Eight states were chosen as the sample for this study. They form the southeast region of the United States Department of Transportation=s National Highway Traffic Safety Administration. The states were Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, and Tennessee. An administrator in each of the sample states was identified utilizing resources from the Motorcycle Safety Foundation and National Association of State Motorcycle Safety Administrators. Information such as name, title, address, and telephone number was verified by contacting the appropriate state office.

Program administrators from the sample states responsible for the day-to-day operations of the motorcycle safety program were invited by letter to participate. The letter requested an interview to be held in her or his state at a location such as an office or at a reasonable off-site location of the administrator=s choice. The letter identified the purpose of the study, the researcher=s motives and study procedures, how the administrator was identified, and potential risks. It was emphasized that the interview was to be completely voluntary, that any questions she or he had would be answered, that identities and state-specific information would be kept anonymous through the use of pseudonyms, and that withdrawal could occur at any time. The open-ended interviews lasted between one and two hours, and were taped and transcribed.

The same primary questions were asked of all administrators. Probing questions varied depending on answers and the need for clarification. The primary questions were piloted by interviewing two motorcycle safety program administrators from outside the sample. This was to assure that the primary questions were appropriately understood. The interview questions were mailed to each program administrator in the sample two weeks prior to the face-to-face interviews.

The interview questions were formulated using the Context-Input-Process-Product (CIPP) model for educational program evaluation, as it was deemed to have characteristics appropriate for addressing program evaluation criteria in motorcycle safety education and training.

The primary questions for each interview were based on the two objectives of this study and on the framework of the CIPP model of evaluation. *Context* evaluation questions were: “How did your program come into being?” “How is your program situated in the state system?” “To what extent was a needs assessment conducted for your program?” “What is the administrative structure of your program?” “How do you evaluate the administrative structure of your program?

Input evaluation questions were: AWhat are the primary objectives of your program?@ "Do you think your program is working up to its capabilities?" "Are there any particular addition you would like to add to your program?" "How do you evaluate the resources you use to achieve program goals?"

Process evaluation questions were: AAre there any design features that prevent your program from being more effective?@ AWhat percent of your time is devoted to program evaluation activities?@ AIf you could change the way the program operates, what would be some of your recommendations?@ AHow do you presently evaluate your program?@ AHow do you evaluate the administrative aspects of your program?@

Product evaluation questions were: AIn what ways should your program be held accountable?@ AWhat outcomes do you measure?@ AWhat would you consider the ultimate measure of accountability for the effectiveness and efficiency of your program?@ AHow do you evaluate the documentation you utilize to determine accountability?@

Answers from program administrators were analyzed as data was collected, and again as the transcripts were reviewed. Evaluation criteria were identified. A two-probe Delphi technique was then implemented to gain further feedback from each program administrator. An alphabetized list of criteria was mailed to each administrator in the sample asking for each criterion to be rated for its value (more value/of value/less value) as characterized by accuracy, clarity and utility. Additionally, the administrators were asked which criteria were presently being used for program evaluation.

The first research question was "What are the current program evaluation practices of state motorcycle safety administrators in their rider education and training program?" This was answered by tallying the results of the first probe. The second research question was "What would state motorcycle safety administrators recommend as ideal program evaluation practices for a rider education and training program?" This was answered by first identifying the criteria rated as having "more value" by at least half of the program administrators, and then conducting a second probe. A list of the remaining criteria was mailed to each administrator again asking them to rate each criterion's value. Those criteria rated by more than half the administrators as having "more value" were considered the criteria for ideal program evaluation practice.

Related Literature

The review of the literature is divided into three parts. Part One reports literature regarding the history and development of educational program evaluation. Part Two provides a summary of scientific studies specific to evaluation of motorcycle safety programs. Part Three provides a review of several program evaluation models.

Part One. **History and Development of Program Evaluation**

The history and development of program evaluation can be viewed from various perspectives. Its enigmatic, complicated nature is exhibited in educational literature. That evaluation itself has many perspectives can be seen in this description in Madaus, Scriven, and Stufflebeam (1983):

> Attempting to evaluate something involves the evaluator coming to grips with a number of abstract concepts, such as value, merit, worth, growth, criteria, standards, objectives, needs, norms, client, audience, validity, reliability, objectivity, practical significance, accountability, improvement, process, product formative, summative, costs, impact, information, credibility, and–of course–with the term *evaluation* itself (p. xi).

Defining evaluation is problematic. Guba and Lincoln (1989) state: "For we argue that there is no 'right' way to define *evaluation*, a way that, if it could be found, would forever put an end to the argumentation about how evaluation is to proceed and what its purposes are. We take definitions of evaluation to be human constructions, whose correspondence to some 'reality' *is not* and *cannot* be an issue. There is no answer to the question, 'But what is evaluation really?' and there is no point in asking it" (p. 21).

One of the problems in discussing evaluation is the need to treat it as a single entity when in actual practice it is a term used to describe several different processes for several different purposes. "Nothing in all the practice of educators in general and of adult education in particular has produced more feelings of guilt, inadequacy, and frustration than evaluation. But, to put it as bluntly as I know how, *I think that evaluation has become a much overemphasized sacred cow*. Furthermore, I think that this very overemphasis has caused an underproduction of practical, feasible, and artistic evaluation in terms of program review and improvement" (Knowles, 1980, p. 198).

Borich and Jemelka (1982) note two areas that inhibit the identification of a clear, precise, and distinct definition of evaluation: "The lack of an adequate theoretical base for the discipline of evaluation has often been cited as a factor that has stifled the development of the field and its ability to provide meaningful evaluative data to practitioners. Even more problematic, however, was the lack of consensus among evaluators as to how evaluations should be conducted" (p. 7).

Smith (1981) identified that the young field of program evaluation was in transition: "Evaluation is being recognized as a highly complex technical, economic, political, and social activity which requires the skills of many professionals – lawyers, economists, artists, scientists, and many others" (p. 7). The Northwest Regional Educational Laboratory initiated the Research on Evaluation Program in 1977. Its purpose was to help devise new methods of educational evaluation through adaptation of metaphorical paradigms and techniques from other disciplines. The laboratory toiled to find new types of evaluation criteria, such as determining to what extent education is democratic, helpful, humane, wholesome, enjoyable, fulfilling, and reflective of highest American

values. "There is general agreement among evaluation theorists and practitioners, both traditionalists and revisionists, that the field of evaluation has not yet fulfilled its promise, not yet lived up to its social role as the provider of relevant, useful, timely information for the assessment of educational and social programs for the establishment of social policy" (Smith, 1981, p. 23). The laboratory made efforts to identify new methods for use in evaluation by studying existing procedures used in seven other fields: investigative reporting, law, architecture, geography, philosophy, literary and film criticism, and watercolor painting. The purposes were to share attempts to use other fields as metaphors for evaluation in order to uncover new evaluation methods, and to encourage a stimulating approach to methodological improvement.

"The growth of program evaluation as a discipline can be linked to the commitment to use public money to create programs for alleviating social, health, and educational problems" (Kosecoff and Fink, 1982, p. 19). The government and citizens alike called for systematic evaluations of the merits of funded programs. The 1960s brought a plethora of publications regarding program evaluation. Noteworthy are Campbell and Stanley's work regarding experimental and quasi-experimental research, which became a cornerstone for evaluation practice, and Scriven's introduction of the concepts of formative and summative evaluation.

Program evaluation models became a vehicle with many uses. Stake (1983) notes: "People expect evaluation to accomplish many different purposes: to document events, to record student change, to detect institutional vitality, to place blame for troubles, to aid administrative decision making, to facilitate corrective action, to increase our understanding of teaching and learning" (p. 29). A primary tension in program evaluation lies in determining its focus, whether to prove accountability or to foster improvement. "Accountability emphasizes looking back in order to assign praise or blame; evaluation is better used to understand events and processes for the sake of guiding future activities" (Cronbach and associates, 1980, p. 4).

Although evaluation activities have been traced back to proficiency requirements for public officials as early as 2200 B.C. (Guba and Lincoln, 1981), Madaus, Stufflebeam, and Scriven (1983) view program evaluation as having six periods of development. They call these the age of reform, the age of efficiency and testing, the Tylerian age, the age of innocence, the age of expansion, and the age of professionalism.

The age of reform, 1800-1900, was a period of economic and technological growth associated with the Industrial Revolution. Educational and social programs scrambled to meet the demands of change and accountability.

The age of efficiency and testing, 1900-1930, was characterized by new scientific management principles that emphasized systemization, standardization, and efficiency.

The Tylerian age, 1930-1945, was associated with Ralph W. Tyler who is considered the father of educational evaluation. The theme of this age was characterized by the philosophy of pragmatism and the methods of behavioristic psychology, and led to the

formulation of measuring specific outcomes. Evaluation was conceptualized as a comparison of intended outcomes with actual outcomes.

Probably the most significant period was the age of innocence from 1946-1957, as it was characterized by an accumulation of data to justify the expansion of existing programs. National, standardized testing became prevalent, but results were not used to judge programs or improve the results of existing structures.

In the age of expansion, 1958-1972, evaluation developed as an industry and emerged as a profession. With the launch of Sputnik in 1957, a strong national interest in the quality of education was born. The value of existing forms of evaluation was scrutinized. Methods to evaluate did not seem to help curriculum developers or those with an interest in program effectiveness. Cronbach (1963) looked at the evaluation efforts of the recent past, and criticized the guiding conceptualizations for their lack of relevance and utility. He advised to turn away from the penchant for post hoc evaluations based on comparisons of the norm-referenced test scores of experimental and control groups, and to couch evaluation in terms of guiding curriculum development. "Hopefully, evaluation studies will go beyond reporting on this or that course and help us to understand educational learning" (Cronbach, 1963, p. 675). Educators were required to shift their thinking from evaluation theory to practice and implementation. This led to a call for new theories and methods of evaluation, as well as for new training programs for evaluators. New models recognized the need to evaluate goals, look at inputs, examine implementation and delivery of services, as well as measure intended and unintended outcomes.

Stufflebeam and Associates (1971) believed that the field of evaluation had been seized by an illness, and they suggested eight symptoms. The symptoms are 1) avoidance, 2) anxiety, 3) immobilization, 4) skepticism, 5) lack of guidelines, 6) misadvice, 7) no-significance-difference, and 8) missing-elements. The avoidance symptom refers to the fact that evaluation processes are considered long, arduous, and complicated processes. The anxiety symptom refers to evaluation perceived as a judgment where program personnel are subjected to an ambiguous process that might possibly yield negative results. The immobilization symptom emerges because evaluations are usually conducted by outside agencies, with little or no internal, ongoing involvement. The skepticism symptom means that the value of evaluation is suspect because experts cannot agree on best practices for obtaining valid results. The lack-of-guidelines symptom means that there is a lack of meaningful, operational guidelines, and agencies that require evaluations cannot provide implementation guidelines. The misadvice symptom stems from the fact that its been shown that experts in the field of evaluation are unable to design or meet criteria of technical soundness. The no-significant-difference symptom means that too often an evaluation technique produces findings that fail to match common observation, and too often evaluation practices fail to discover significant differences. The missing-elements symptom referred to the lack of adequate theory, the lack of specifications for usefulness, the lack of appropriate instruments and design, the lack of mechanism for organizing, processing, and reporting, and the lack of trained personnel.

Borich and Jemelka (1982) noted that several items were added to the above symptoms. "And to these were added the lack of trained personnel, the lack of knowledge about decision processes, the lack of values and criteria for judging evaluation results, the need to have different evaluation approaches for different types of audiences, and the lack of techniques and mechanisms for organizing, procuring, and reporting evaluative information" (p. 6).

The final period, from 1973 to the present is considered the Age of Professionalization. This period began with the field in disarray. "Evaluation studies were fraught with confusion, anxiety, and animosity. Evaluation as a field had little stature and no political clout" (Madaus, Scriven, and Stufflebeam, 1983, p. 15). As the field moved into meta-evaluation processes and as professional preparation programs were implemented, the body of knowledge moved from occasional papers and periodical publications to standardization of methods. "During this period, evaluators increasingly realized that the techniques of evaluation must achieve results previously seen as peripheral to serious research; serve the information needs of the clients of evaluation; address the central value issues; deal with situational realities; meet the requirements of probity; and satisfy needs for veracity" (Madaus, Scriven, and Stufflebeam, 1983, p. 16). Program evaluation was perceived to be an immature profession, and it was during this time period that several models of program evaluation were developed. "Ultimately the value of program evaluation must be judged in terms of its actual and potential contributions to improving learning, teaching and administration, health care and health, and in general the quality of life in our society" (Madaus, Scriven, and Stufflebeam, 1983, p. 18).

Presently the field of evaluation is diffuse. Educational programming is often deeply contextual and, if evaluated, requires idiosyncratic evaluation methodologies. Guba and Lincoln (1989) view evaluation in terms of three generations, and offer an approach they call fourth generation evaluation. "It is our intention to define an emergent but mature approach to evaluation that moves beyond mere science--just getting the facts--to include the myriad of human, political, social, cultural, and contextual elements that are involved. We have called this new approach *fourth generation evaluation* to signal our construction that this form moves beyond previously existing generations, characterized as measurement-oriented, description-oriented, and judgment-oriented, to a new level whose key dynamic is *negotiation*" (p. 8).

They describe the first generation of evaluation as measurement. It had its basis in teaching and evaluating school children. Measured was the ability to regurgitate facts and truths. Another impetus for this type of evaluation occurred because of the need for armed service personnel in World War I. Also, applying scientific measurement in the social sciences became prevalent, and the movement was rampant in industrial environments. "Psychology in particular became wedded to the new scientific approach, attempting to emulate the physical sciences as closely as possible" (Guba and Lincoln, 1989, p. 25). The role of the evaluator was technical, as seen in the multitude of school tests in the 1920s and 1930s.

Second generation evaluation was characterized by discovering if programs were effective. Ralph W. Tyler, a member of the Bureau of Educational Research at Ohio State University, was developing tests that would measure whether or not students were learning what their professors had intended. These learning outcomes were labeled objectives. "Tyler was engaged to carry out the same kind of work with the Eight Year Study secondary schools, but with one important variation from conventional evaluation (measurement): the purpose of the studies would be to *refine the developing curricula* and *make sure they were working*. Program evaluation was born" (Guba and Lincoln, 1989, p. 28).

Third generation evaluation was characterized by judgment. The launch of Sputnik exposed some serious flaws with simply measuring the attainment of objectives. "The call to include judgment in the act of evaluation marked the emergence of third generation evaluation, a generation in which evaluation was characterized by effort to reach *judgments*, and in which the evaluator assumed the role of *judge*, while retaining the earlier technical and descriptive functions as well" (Guba and Lincoln, 1989, p. 30). Models were developed that called for the evaluator to also be a judge.

None of the first three generation of program evaluation met the growing demands for utility. Guba and Lincoln (1989) identified three flaws: a tendency toward managerialism (in which the clients or sponsors that commissioned the evaluation have too much control), a failure to accommodate value-pluralism (whose values are to dominate the evaluation), and overcommitment to the scientific paradigm of inquiry (omitting the contextual richness of evaluation with an overdependence on quantitative measurement).

Fourth generation evaluation, responsive constructivist evaluation, is a form of evaluation in which claims, concerns, and issues of stakeholders serve as organizing factors. It uses as organizers the claims, concerns, and issues about that which is to be evaluated. A claim is any assertion that a stakeholder may introduce that is favorable to the evaluation. A concern is any assertion that a stakeholder may introduce that is unfavorable to the evaluation. An issue is any state of affairs about which reasonable persons may disagree. Three broad classes of stakeholders are identified: the agents, those persons involved in producing, using, and implementing the evaluation; the beneficiaries, those persons who profit in some way from the use of the evaluation; and the victims, those persons who are generally affected by the use of the evaluation.

Meyers (1981) offers another perspective regarding the motivations for conducting program evaluation, particularly in the realm of social programs intended to ameliorate social problems. He asks the question: "Is there a need for program evaluation?" He believes the changes in recent times reflect political developments rather than advances in evaluation methodology. He goes on to suggest four motivations for interest in program evaluation, beginning with a direct connection to the highest office in the land. "Taking office in 1969, the Nixon administration wanted to cut programs, partly in revenge against the Democrats and partly to save money. The new administration immediately showed a strong interest in evaluation, since it appeared to be a way to legitimately

abolish programs. They believed that most social programs do not work and that logical analyses of costs and benefits would demonstrate that fact (Meyers, 1981, p. 2).

A second motivation relates to the planning-programming-budgeting system (PPBS). Its basis lies in the business world, and it looks for results through cost-analysis and cost-benefit approaches. "The data needed for the cost-benefit analyses required the evaluation of outcomes of programs; indeed a major problem in implementing PPBS was that the necessary data were not available (Meyers, 1981, p. 2).

A third motivation is tied to accountability. "The taxpayers' revolt, the general antibureaucratic sentiment, and the traditional American wish to be free of external controls, fuel criticism of public bureaucracies that fund social programs" (Meyers, 1981, p. 3). He goes on to say that those demanding program evaluation rely on two facts: that outcome evaluations show most programs do not work, and that administrative costs approach the values of services provided to program recipients.

A fourth and final motivation is the conservative movement, which includes "an antiintellectual, antiuniversity, anti-Ivy League, and antiliberal animus. Evaluation is one way to apply conservative, businesslike realism in order to abolish programs" (Meyers, 1981, p. 3).

Baugher (1981) provides recommendations appropriate for developing a successful measurement program. He identifies ten areas for implementers of evaluation strategies to consider. These include the following: determining which type of effectiveness is of greatest concern, as there is no single model of effectiveness; use of multiple indicators of efficacy, as varying perspectives should be honored; emphasizing the importance of effectiveness measurement, as long-term survival is important; focusing on solutions and not problems, as addressing mistakes leads to defensiveness; planning for evaluation, as evaluation can be more comprehensive; making evaluation an ongoing process, as this can lead to efficiency in effecting change; accurate documentation, as results can have serious consequences; use of appropriate methodological approaches, as this will provide the best look at changes that have occurred; careful communication, as clarity will minimize misuse of the evaluation information; and the need for replication studies, as this adds greater strength to the conclusions.

Fetterman (1995) in responding to comments about objectivist evaluations stated: "I understand and appreciate this idealistic view of reality; however, anyone who has recently had to roll up their sleeves and get their hands dirty in program evaluation or policy arenas is aware that evaluation, like any other dimension of life, is political, social, cultural, and economic. It rarely produces a single truth or conclusion" (p. 189).

Cronbach and associates (1980) argue for a comprehensive transformation of program evaluation. They urge that program evaluation not have as a purpose to eliminate the fallibility of authority or bolstering its credibility, but rather to facilitate a democratic, pluralistic process by enlightening all participants. "Evaluation has vital work to do, yet its institutions and its ruling conceptions are inadequate. Enamored of a vision that

'right' decisions can replace political agreements, some who commission evaluations set evaluators on unrealistic quests" (Cronbach and associates, 1980, p. 1). They provide 95 theses as principal points to spur debate regarding the fundamental issues surrounding program evaluation.

Two recent publications provide program evaluation ideas for traffic safety practitioners, which would include motorcycle safety administrators. They are *Demonstrating Your Program's Worth: A Primer on Evaluation for Programs To Prevent Unintentional Injury* (Thompson and McClintock, 1998), and *The Art of Appropriate Evaluation: A Guide for Highway Safety Program Managers* (National Highway Traffic Safety Administration, 1999).

The former booklet is designed to show program managers how to demonstrate the value of their work to the public, to their peers, to funding agencies, and to the people they serve. It suggests formative evaluation during the development of a new program to maximize the likelihood that the program will succeed, process evaluation as soon as a program begins operation to identify early any problems that occur in reaching the target population, impact evaluation to move resources from nonproductive to productive areas, and outcome evaluation to determine the degree to which a program has met its ultimate goals.

The latter booklet is designed to provide an overview of the steps that are involved in program evaluation. It acknowledges that ideally a direct cause and effect relationship should be determined between the countermeasure and results. Suggested are solid research designs with random assignment to experimental and control groups and sophisticated statistical analysis. But exception is taken in regard to traffic safety evaluation when it is stated: "Traffic safety evaluation is an applied science that works within the constraints of state and local program implementation. Most local communities simply do not have the volume of traffic deaths and injuries to conduct that kind of countermeasure effectiveness of evaluation" (National Highway Traffic Safety Administration, 1999, p. 10-11).

Boulmetis and Dutwin (2000) provide practical techniques aimed at improving the competencies of program and project managers who have little experience with program evaluation. "Both the person who is designing and performing an evaluation and the person who is participating in and receiving the findings of an evaluation need to prepare themselves to understand the basic processes involved" (p. x.). Common concepts proposed include efficiency (the degree to which a program or project has been productive in relationship to its resources), effectiveness (the degree to which goals have been achieved), and impact (the degree to which a program or project resulted in changes). The text also makes a distinction between evaluating to see if objectives have been achieved and evaluating in order to make a decision.

Part Two. **Evaluation of Motorcycle Safety Programs**

Part Two provides a summary of formal research related to motorcycle safety programs. A technical paper, published by the National Highway Traffic Safety Administration, supported the development of programs to assure that motorcycle operators have the skill and knowledge to operate a motorcycle safely in traffic (Buchanan and Tarrants, 1982). The impact of such programs in reducing crashes is unknown. "Surprisingly few evaluations have been undertaken to determine the effectiveness of motorcycle rider education/training despite the importance ascribed to these programs in most jurisdictions" (Mayhew and Simpson, 1996, pp. 29-30). Winn and McPherson (1990) surveyed several states that passed programs between 1979 and 1988, and found that few required routine evaluations or improvements in the collection of fundamental motorcycle-safety-data elements. "Results showed that even relatively inexpensive administrative-cost and demographic evaluations are not being conducted and almost no impact evaluations are being undertaken" (p. 6-98). They recommended that a set of motorcycle-program- specific evaluation criteria be established and tested in a small sample of states to assist and guide new and existing programs, and that states considering new programs should consider evaluation as an integral part of the legislation.

Only a handful of formal research studies exist relative to program evaluation for motorcycle safety programs. These range from self-report surveys to experimental research designs.

Higham (1980) reported low crash rates for graduates of U.S. Air Force motorcycle safety courses. Over three years after the introduction of the course there had been on average a 40 percent reduction in motorcycle crashes. But Lonero (1998) noted: "Without controlled experiments, of course, it is not clear that either of these courses caused the apparent reduction in new riders' crashes" (p. 50).

Osga (1980) used self-reported information on formally trained and untrained riders in South Dakota. Course graduates had a higher accident rate for mileage covered before and after the course than the untrained subjects. It was concluded that MSF Rider Course graduates are as likely to become involved in accidents as untrained riders.

Jonah and others (1982) evaluated the Canada Safety Council's Motorcycle Training Program (MTP). Surveys found no significant differences between MTP-trained and untrained riders in the rate of reported crashes when age, sex, education, licensing age, exposure, and riding after drinking were controlled. However MTP graduates had fewer moving violations. Self-selection may have been the issue. Of note is that official driving records underestimated the incidence of crashes as measured by self-report.

Mortimer (1984) compared self-reported collision experiences of course graduates with those not completing a course in Illinois, and found that when controlling for age and years licensed, course graduates did not have a lower accident rate. Since the study was

retrospective, there was no random assignment of persons to the experimental or control groups as a means of eliminating bias.

An experimental study, sponsored by the National Highway Traffic Safety Administration, was completed in the state of New York in the early 1980s. Mayhew and Simpson (1996) reviewed the results, and report no statistically significant differences in crash rates as a function of different training and licensing systems. McPherson (1989) notes that the study was plagued by operational problems. "The applicant flow was considerably less than had been expected and administrative problems led to long delays between dates of application, training, and license testing. Also, the applicants utilized in this study were found not to be novices, having an average of three years' experience" (p. 7).

McKnight (1987) studied the motorcycle safety program in the state of Pennsylvania. More than 3,000 pairs of trained and untrained riders license applicants were matched on the basis of age, sex, and prior driving records. There were no significant differences between the mileage-adjusted collision rates of trained and untrained riders.

Rothe and Cooper (1987) after reviewing evaluations of motorcycle rider education programs commented: "Most motorcycle rider education program evaluations were designed to judge the effectiveness of the program's impact on trained motorcyclists' driving behavior. The criteria for measurement were motorcycle accidents and violations" (p. 27). They recommend obtaining documentation of how motorcycle education is delivered before defining what course elements may produce specific outcome behavior.

Mortimer (1988) evaluated the Motorcycle Safety Foundation's Motorcycle Rider Course (MRC) in Illinois. Training did not reduce self-reported crashes when age and years licensed were controlled, nor did it affect violation frequency or cost of damage. Training was found to be associated with use of safety gear and with lower medical cost per crash. Based on exposure, trained riders had more crashes. Many course participants do not continue to drive motorcycles, so the course may allow people to learn to ride a motorcycle in a safe environment, and then decide not to ride on the road.

McDavid and others (1989) reported that a five-year study completed in 1989 regarding the British Columbia Safety Council motorcycle program found prior differences between people who take motorcycle training and those who do not. In matching trained and untrained riders using motor vehicle records, the untrained group had 64 percent more motorcycle crashes but only 24 percent more non-motorcycle crashes. The effect was strongest in the first year of training. The authors suggested that the training made a difference, although the study design did not permit a clear causal difference. Even though subjects were matched, self-selection into training occurred and could account for the differences. Simpson and Mayhew (1990) pointed out that the subjects were not matched on riding exposure, which may be the most critical variable.

Simpson and Mayhew (1990) reviewed the literature on motorcycle rider training and similarly found no evidence of crash prevention. They note that although trained riders may have higher skill levels, it is reasonably clear that skill is not the most critical factor in crashes. "Such initiatives are founded on the fundamental and compelling assumptions that students who are exposed to the education/training will be at lower risk of traffic mishap than those who are not. Somewhat surprisingly, however, research has not been able to establish the loss-reduction value of formal motorcycle education/training courses" (p. 258). They point out the need for motivational methods, such as linking safety with health promotion and the development of community-oriented controls. Billheimer (1996) studied the California Motorcyclist Safety Program (CMSP). Pre-post comparisons of collision trends revealed that the total number of motorcycle collisions and fatal motorcycle collisions had decreased since the implementation of the program. It is unclear which factors, such as changing demographics, fewer younger riders, or the number of licensed riders, could have accounted for the results. Simpson and Mayhew (1996) note: "Given such methodological concerns, it must be concluded that the study provides little in the way of conclusive evidence that the CMSP contributed to the overall crash reductions witnessed in California from 1987 to 1995, or that it reduced the crash involvement of novice riders during the six months following training" (p. 34).

Commenting on their review of several studies of motorcycle safety programs, Mayhew and Simpson (1996, p. 34) state: "In summary, several studies have failed to provide definitive conclusions about the effectiveness of rider education and training in reducing crashes."

Part Three. **Models of Program Evaluation**

Part Three presents a synopsis of ten models of program evaluation that are found in literature regarding educational programs. These models are presented to show the array of models that have been proposed for a variety of educational contexts. Models have limitations. "Although evaluation models usually have little relationship to perspectives or needs, more than fifty models of program evaluation have emerged and have gained some acceptance" (Steele, 1991, p. 262). Models are not necessarily templates or procedures to follow, but they can provide the basis for developing a plan specific to a particular program. Also, models differ in their intent, as most are designed for program improvement. This is clarified by Boulmetis and Dutwin (2000): "All the models for evaluation differ from research strategies in that evaluation results are provided to the appropriate stakeholders for the purpose of program improvement. The purpose of research, in contrast, is to draw causal links between observed phenomena and to add to the knowledge base on those phenomena, and the audience in the professional field in general" (p. 69). Also, it must be acknowledged that models are sometimes abstract and are often modified for particular contexts. "Indeed, from one standpoint the role of a good model is to speed its own obsolescence. It cannot provide final answers and is not intended to. It has served its purpose if it provides fresh insights into the working of things" (Pfeiffer, 1968, p. 27).

Tyler Model

Program evaluation can be closely associated with specific objectives, whether they are related to individual students or whether they are related to curricula and programs. Tyler's evaluation process is directly tied to the concept of curriculum objectives. He states: "The process of evaluation is essentially the process of determining to what extent the educational objectives are actually being realized by the program of curriculum and instruction" (Tyler, 1949, p. 69). It is his view that evaluation is the process for determining the degree to which changes in behavior are occurring. "Until this time evaluation had existed largely for the purpose of making judgments about individual students in relation to test norms and of labeling the students as overachievers, underachievers, or 'normal' achievers" (Guba and Lincoln, 1981, p. 5).

Considered an educational psychology model, Tyler's work was influential in other evaluation models of that era. Generally these models consisted of the following steps: 1) identify objectives, 2) state objectives in measurable behavioral terms, 3) devise and administer measurement outcomes, 4) compare obtained results with the objectives that were specified (Borich and Jemelka, 1982, p. 7). Such models were used extensively to discover information regarding program achievement of defined objectives, and to make adjustments with refinements and revisions, a process that would later be called formative evaluation. Results were not available until after the study rather than during a study.

The Tyler model was based on a scientific pre-post paradigm. "Tyler's insistence that a behavior needed to be measured twice—before and after the "treatment" afforded by the curriculum—made the rationale a 'natural' for the usual experimental design approach espoused in other behavioral science areas, for example, psychology" (Guba and Lincoln, (1981, p. 7).

Stake Countenance Model

The Countenance Model involves completing a description matrix and a judgment matrix (description and judgment being the two countenances). "To be fully understood, the educational program must be fully described and fully judged" (Stake, 1967, p. 525.) Each matrix is divided into two columns. The description matrix consists of intents and observations; the judgment matrix consists of standards and judgments. Both matrices are divided into three rows labeled antecedents, transactions, and outcomes. The task for a program evaluator is to determine the intents at all three levels, collect data for the observations, and interpret discrepancies between observed performance and standards (Guba and Lincoln, 1981, p. 12).

Advantages of the Countenance Model are that context is considered in the evaluation, and judgment is an integral component of the model. Disadvantages are that a method for determination of the standards was not specified, and a way to capture unintended effects was overlooked (Guba and Lincoln, 1989, p. 13-14). "Although he explicitly warned the evaluator not to overlook unintended effects, Stake failed to provide guidance

on how to find and take account of them. He continued with and emphasis on formal evaluation, and this emphasis tied evaluation even more closely to the scientific paradigm and its attendant measurement processes" (Guba and Lincoln, 1981, p. 14). Additionally, the model was quite complex and practitioners had difficulty in comprehending it.

The Countenance Model is considered a refinement of the Tyler approach as it encourages the examination of relationships to the process. This involves examining the relationships among antecedents, transactions and outcomes, determining congruency, and making judgments about the strengths and weaknesses of a program. "Illogical contingencies constituted possible program weaknesses" (Borich and Jemelka, 1982, p. 7).

Scriven Goal-free Model

In goal-free evaluation program goals are not criteria on which the evaluation is based. "Instead, the evaluation examines how and what the program is doing to address needs in the client population" (Boulmetis and Dutwin, 2000, p. 73). This model grew out of the need to move beyond objective-based evaluations so that decisions made by program principals could be taken into account over program objectives, the need to have the evaluation focus not only on results but on ways to refine and improve program processes, and the need to maximize the utility of an evaluation. Course performance could be considered more important than a comparative analysis. Scriven (1967) suggests that evaluators not consider program objectives when conducting an evaluation. He was puzzled as to why intended and unintended effects were separated. "Hence, Scriven came to the conclusion that evaluation should be *goal free*, that is, that it should evaluate actual effects against a profile of demonstrated needs in education. Thus, Scriven's organizer became *effects* rather than goals or decisions" (Guba and Lincoln, 1981, p. 17).

Goal-free evaluation meant that a favorable evaluation could be a result simply by demonstrating that a product or service was responsive to a need. "Scriven has been primarily concerned with reducing the effects of bias in evaluation. This model reduces the bias of searching only for the program developers' prespecified intents by not informing the evaluator of them. Hence, the evaluator must search for all outcomes" (House, 1983, p. 46). An opposing view is offered by Meyers (1981) as he states: "Interaction with program staff may be conducive to certain biases in the evaluator, but lack of contact with the program staff is conducive to other biases. Consequently, the claim that the goal-free evaluation method abrogates bias must be rejected" (p. 123). It may be that goal-free evaluation has little merit, as he adds: "In fact, the idea that one could look at a program without readily perceiving its goals is itself unrealistic; the content of the program and the choice of pre- and post measures certainly reveal information about the goals. It is to be doubted, then, whether goal-free evaluation is a sound way of avoiding the issues posed by program goals" (p. 126).

Eisner Connoisseurship Model

The Connoisseurship model relies on a human being as a measurement instrument. Eisner (1985) connects program evaluation with art criticism; he states: "To achieve such ends, an educational critic must possess a high level of connoisseurship within the area that he or she criticizes. Connoisseurship is the art of appreciation, and criticism is the art of disclosure" (p. 237). Data collection, analysis, processing, and interpretation take place within the mind of the evaluator as judge, and hence are not open to direct inspection. The appropriateness of this is displayed in Eisner's philosophy of classroom instruction. "Because I believe teaching in classrooms is ideographic in character, that is, because I believe the features of classroom life are not likely to be explained or controlled by behavioral laws, I conceive the major contribution of evaluation to be a heightened awareness of the qualities of that life so that teachers and students can become more intelligent within it" (Eisner, 1983, p. 339). This philosophy may also be appropriate in program evaluation.

The value of the Connoisseurship Model is significant as it opened the evaluation door to an entirely different way to approach program evaluation. A transition from the positivistic, quantitative way of knowing gave way to a constructivist, qualitative line of thought. "It is in effect a nonscientific supplement to traditional evaluation models, and it demonstrates that the scientific paradigm is not essential to the development of a powerful and useful evaluation approach. The connoisseurship model has the honor of being the first to break cleanly with that paradigm" (Guba and Lincoln, 1981, p. 20).

Stufflebeam and Webster (1983) point out the primary advantage and disadvantage. They state: "The main advantage of the connoisseur-based study is that it exploits the particular expertise and finely developed insights of persons who have devoted much time and effort to the study of a precise area. They can provide an array of detailed information that the audience can then use to form more insightful analysis than otherwise might be possible. The disadvantage of the this approach is that it is dependent on the expertise and qualifications of the particular expert doing the evaluation, leaving much room for subjectivity, bias, and corruption" (p. 35).

Stake Responsive Model

The genesis for the Responsive Model is Stake's perspective of evaluation regarding educational programs. "Evaluation is an *observed value* compared to some *standard*. It is a simple ratio, but this numerator is not simple. In program evaluation, it pertains to the whole constellation of values held for the program. And the denominator is not simple, for it pertains to the complex of expectations and criteria that different people have for such a program" (Stake, 1983, p. 291). He does not see the role of an evaluator as one of solving equations, but rather one of making a comprehensive statement of what the program is observed to be, with useful references to the satisfaction and dissatisfaction of selected people.

Responsive evaluation is based on what people do naturally to evaluate things. "An educational evaluation is responsive evaluation if it orients more directly to program activities than to program intents, if it responds to audience requirements for information, and if the different value perspectives of the people at hand are referred to in reporting the success and failure of the program" (Stake, 1983, p. 293). This type of evaluation focuses on the issues as opposed to objectives in order to reflect the complexity, immediacy, and value of a program.

"Responsive evaluation is an emergent form of evaluation that takes as its organizer the *concerns and issues of stakeholding audiences*" (Guba and Lincoln, 1981, p. 23). Stakeholders are persons or groups that are put at some risk by an evaluation. The concerns and issues are gathered in conversations with persons in and around the program. Stake's suggestions are noted in Guba and Lincoln (1981, p. 25-26) as 12 interactive steps. Generally, these entail talking with stakeholders, making observations, determining the purposes of project and stakeholder concerns that include issues and problems that the evaluation should address, design an evaluative structure with human instruments, collect data, and organize a reporting structure qualitatively and/or quantitatively.

Because stakeholders may have differing constructions regarding the value of a program, multiple responses need to be organized to effectively communicate evaluation results toward reaching a consensus. "… one of the major tasks of the evaluator is to conduct the evaluation in such a way that each group must confront and deal with the constructions of all the others, a process we shall refer to as a hermeneutic dialectic" (Guba and Lincoln, 1989, p. 41).

Guba and Lincoln Responsive Constructivist Model

Naturalistic inquiry is a paradigm of inquiry; that is, a pattern or model for how inquiry may be conducted. It is characterized by discovery in natural settings, typically uses a case study format, and leans toward qualitative methods. Guba and Lincoln (1983, pp. 315-323) provide five axiomatic differences between what they call the rationalistic paradigm or scientific method, and the naturalistic paradigm. These axiomatic differences include the nature of reality, the nature of the inquirer-object (or respondent) relationship, the nature of truth statements, assumptions about causal relationships, and the role of values within disciplined inquiry.

Guba and Lincoln (1983) name six characteristic postures, which they call a synergistic set, that distinguish naturalistic inquiry. These are the preferred methods, the source of theory, the knowledge types used, the instruments, the design, and the setting (pp. 323-325). Preferred methods refers to the fact that quantitative methods have greater precision and are mathematically manipulable, while qualitative methods are richer and can deal with phenomena not readily translatable into numbers. The theories preferred by rationalists are *a priori*, while naturalists prefer theories to arise from the data rather than being imposed on them. Knowledge type refers to the rationalist's preference toward explicit language while the naturalist builds upon tacit knowledge. The rationalist prefers

non-human data collection instruments; the naturalist prefers humans as instruments. The rationalist utilizes a preordinate design while the naturalist prefers to see the design emerge as the inquiry proceeds. The laboratory is the preferred setting for rationalistic study; a natural setting is preferred for naturalistic study, which is the normal setting and situation inherent in a program.

Guba and Lincoln (1989, pp. 252-269) further developed their views regarding naturalistic inquiry into what they call responsive, constructivist evaluation. They use the term responsive to designate a different way of focusing an evaluation, and the term constructivist to designate the methodology actually employed. Calling this line of inquiry a fourth generation view of evaluation, they offer seven distinguishing characteristics: evaluation is a sociopolitical process; evaluation is a joint, collaborative process; evaluation is a teaching/learning process; evaluation is a continuous, recursive, and highly divergent process; evaluation is an emergent process; evaluation is a process with unpredictable outcomes; and evaluation is a process that creates reality.

Provus Discrepancy Model

Provus proposed a five-stage evaluation process defined as design, installation, process, product, and cost-benefit analysis. This consists of documenting a description of the program that includes inputs, processes and outcomes, observing field operations to collect information about the discrepancy between expected and actual operation, relating component parts of the program to short-term, enabling behaviors as displayed by participants, relating component parts of the program to end-of-program behaviors, and comparing an experimental program with a realistic alternative (Borich and Jemelka, 1982, p. 9).

This evaluation process focuses on weaknesses of a program by identifying discrepancies between intended and actual outcomes. Each stage is evaluated and program evaluation does not continue until a decision is made that results are adequate to move on, or that the program standards or operations need to change.

Boulmetis and Dutwin (2000) state: "The model assumes the aim is not to prove cause-and-effect relationships but to understand the evidence well enough to make reasonable assumptions about cause-and-effect" (p. 71). They acknowledge that the discrepancy model may be appropriate in a program where staff and an evaluator can work collaboratively from the outset of program operation.

Kilpatrick Evaluation Model

Kilpatrick (1994) offers a pragmatic model for use in devising program evaluation. His model consists of a four-level approach. It has been embraced by many segments of the human resources training community, although its implications are applicable to educational programs. In differentiating between the terms training and education, Kilpatrick states: "Although a distinction is often made between these two terms, for simplicity I have chosen to speak of them both simply as *training* and to emphasize

courses and programs designed to increase knowledge, improve skills, and change attitudes, whether for present job improvement or development in the future" (p. xiv).

Kilpatrick sees evaluation as consisting of four levels or types: reaction, learning, behavior, and results. Reaction evaluation takes place periodically throughout a program and provides information from the participant perspective. This information can be used to make changes in design, methods, personnel, and facilities. Learning evaluation, often measured by pre-testing and post-testing, provides information about the knowledge, attitude, skills, and values gained by participants. Behavior evaluation provides information about changes in actual performance in real-world environments. Results evaluation provides information regarding return-on-investment or cost-benefit analysis, such as improved quality, increased productivity, and (in the context of safety) lowered accident rates.

Empowerment Evaluation

Empowerment evaluation is an emerging concept that has been adopted in higher education, government, inner-city public education, and foundations throughout the United States and abroad. It is a method for using evaluation concepts, techniques and findings to foster improvement and self-determination (Fetterman, 1996, p. 4). It is highly situational and context specific. There are five facets to empowerment evaluation: 1) training participants to conduct their own evaluation, 2) evaluators as facilitators and coaches rather than judges, 3) evaluators advocating on behalf of groups advocating themselves, 4) illumination, and 5) liberation for those involved. Empowerment evaluation "is designed to help people help themselves and improve their programs using a form of self evaluation and reflection (Fetterman, 1996, p. 5). This approach avoids a narrow paradigm regarding evaluation and recognizes that differing needs and contexts require differing evaluative responses. "The focus should be on the problem or issue; methods and methodologies should follow, not precede. Moreover, a singular approach to evaluation is not responsive to the needs and demands of program participants and clients who live in a rapidly changing, highly unstable environment" (Fetterman, 1995). Although the fluidity of empowerment evaluation that accommodates chaos and ambiguity is counter to a positivistic view, it can provide ways to understand program operations. "It is responsive to rapid and unexpected shifts in program design and operation because it requires continual collection, description, reflection, and feedback on information about a group or organization in all its complexity" (Fetterman, 1996, p. 380).

Stufflebeam CIPP Model

"The CIPP approach is based on the view that the most important purpose of evaluation is not to prove but to improve" (Stufflebeam, 1983, p. 118). The model was developed by the Phi Delta Kappa Commission on Evaluation as a result of attempts to evaluate projects that had been funded through the Elementary and Secondary Education Act of 1965.

The CIPP Model defines evaluation as the process of delineating, obtaining, and providing useful information for judging decision alternatives (Stufflebeam, 1971, p. xxv). Four kinds of decisions are specified. They are: 1) planning decisions to determine objectives, 2) structuring decisions to establish procedural designs, 3) implementation decisions to execute designs and 4) recycling decisions to determine whether to continue, terminate, or modify a project. "Fundamentally, the use of the CIPP Model is intended to promote growth and to help the responsible leadership and staff of an institution systematically to obtain and use feedback so as to excel in meeting important needs, or, at least, to do the best they can with the available resources" (Stufflebeam, 1983, p. 118). The model divides evaluation into four distinct strategies—**C**ontext evaluation, **I**nput evaluation, **P**rocess evaluation, and **P**roduct evaluation, thus the acronym CIPP (Borich and Jemelka, 1982, p. 10).

Context evaluation provides information about the needs, problems, and opportunities in order to identify objectives. "Its purpose is to provide a rationale for determination of objectives" (Stufflebeam, 1971, p. 118). It defines the relevant environment, describes the desired and actual conditions pertaining to that environment, identifies unmet needs and unused opportunities, and diagnoses the problems that prevent needs from being met and opportunities from being used.

Input evaluation provides information about the strengths and weaknesses of alternative strategies for achieving given objectives. "This is accomplished by identifying and assessing (1) relevant capabilities of the responsible agency, (2) strategies for achieving program goals, and (3) designs for implementing a selected strategy" (Stufflebeam and others, 1971, pp. 222-223).

Process evaluation provides information about the strengths and weaknesses of a strategy during implementation so that either the strategy or its implementation might be strengthened. "Process evaluation has three main objectives—the first is to detect or predict defects in the procedural design or its implementation during the implementation stages, the second is to provide information for programmed decisions, and the third is to maintain a record of the procedure as it occurs" (Stufflebeam and others, 1971, p. 229).

Product evaluation provides information for determining whether objectives are being achieved and whether the procedure employed to achieve them should be continued, modified or terminated. "The general method of product evaluation includes devising operational definitions of objectives, measuring criteria associated with the objectives of the activity, comparing these measurements with predetermined absolute or relative standards, and making rational interpretations of the outcomes using the recorded context, input, and process information" (Stufflebeam and others, 1971, p. 232).

Although the CIPP Model was originally developed to provide timely information in a systematic way for decision making in order to provide proactive application of evaluation, it can function as a retroactive purpose of providing information for accountability (Stufflebeam, 1972, p. 3). Characteristics of the model are that it considers evaluation as a systematic, continuing process; that the evaluation process

includes the three basic steps of delineating questions to be answered and information to be obtained, the obtaining of relevant information, and the providing of information to decision makers for their use to make decisions and thereby to improve ongoing programs; and that evaluation serves decision making. "It meshes well with traditional organizational training evaluation and with the value-added training evaluation model" (Philippi, 1996, p. 838).

In an update of the CIPP Model, Stufflebeam (1983, p. 140) provides the following perspective:

> But evaluation is also a necessary concomitant of improvement. We cannot make our programs better unless we know where they are weak and strong and unless we become aware of better means. We cannot be sure that our goals are worthy unless we can match them to the needs of the people they are intended to serve. We cannot plan effectively if we are unaware of options and their relative merits; and we cannot convince our constituents that we have done good work and deserve continued support unless we can show them evidence that we have done what we promised and produced beneficial results. For these and other reasons, public servants must subject their work to competent evaluation, which must help them sort out the good from the bad and point the way to needed improvements.

Brookfield (1986) points out that a benefit of the CIPP model is that it includes contextual scrutiny of a program's origins, implementation, continuing operations, as well as its final achievements. "In any evaluation undertaken according to this model, the influence of institutional priorities, the impact of individual personalities, and the importance of the prevailing political climate would receive due acknowledgement. At present, many evaluations lack this critical contextual component" (p. 270).

Findings

An analysis of the data produced 205 criteria for consideration in program evaluation practice. The first research question was "What are the current program evaluation practices of state motorcycle safety administrators in their rider education and training program? The answer to this question was answered in the first questionnaire in which administrators noted which criteria were currently being utilized in their programs. Of the 205 evaluation criteria, analysis shows that only 17 criteria are utilized by more than half of the program administrators in current evaluation practice. Table 1 provides the program evaluation criteria currently used in evaluation practices. (Appendix A is the list of the 205 criteria and the tallies from the first probe.)

Table 1

Criteria Currently Used for Program Evaluation	
Accomplishment of Predetermined Goals	Frequency of Instructor Updates
Capturing and Acting on Input from Instructors	Having Skill Test Waiver Upon Course Completion
Capturing and Documenting Unsolicited Responses from Participant	Maintenance Condition of the Motorcycles
Course Student Evaluation Results	Numbers Trained
Degree of Assessing Individual Sites, Not State as a Whole	Pass/Fail Rate in Courses
Degree of Formal Documentation of Quality	Quality of Instruction and Instructors
Degree of Professional Development Among Instructors	Quality of Instructor Updates
Dollars Spent on Quality Assurance	Requiring That Student Evaluations Be Completed
	Uniformity of Course Reporting

The second research question was "What would state motorcycle safety administrators recommend as ideal program evaluation practices for a rider education and training program? To answer this question, a second questionnaire was developed and mailed to each administrator that contained the 46 criteria that were named by more than half of the sample as having "more value" on the first questionnaire. Analysis shows that a total of 30 criteria were named by more than half of the program administrators in the second questionnaire as having "more value" in ideal program evaluation practices. (Appendix B is the list of these criteria and tallies from the second probe.) Table 2 provides the program evaluation criteria identified for ideal program evaluation practice.

Table 2

Criteria for Ideal Program Evaluation	
Ability to Remedy/Remove Problems Within Program	Extent to Which Quantity and Quality Are Raised Simultaneously
Amount of Funding	Having Adequate Budget to Meet Expectations
Capturing and Acting On Input from Instructors	Having Adequate Number of Motorcycles
Clear Communication as the "Measuring Stick" for Program	Increased Learning by Students
Course Student Evaluation Results	Increased Skill by Students
Dedication of the People Involved	Level of Communication Within Program
Degree of Emphasis on Service Function	Maintenance Condition of Motorcycles
Degree of Professional Development Among Instructors	People Skills of Administrators and Instructors
Degree of Trust Within Program	Priority of Agency in Which Program Is Conducted
Dollars Spent on Quality Assurance	Quality of Instruction and Instructors
Extent of Constant Learning Within Program	Quality of Instructor Updates
Extent to Which Communication Is Open	Quality of Interaction between Instructors and Participants
Extent to Which Program Participation Is Fun	Support From Superiors
Extent to Which Quality Assurance Measures Are Implemented	Training Itself Is Safe
	Training Site Coverage in State
	Way Problems Are Addressed

Seven criteria, shown in Table 3, were on both the list of criteria currently used and the list of criteria for ideal practice. These include: 1) Capturing and Acting on Input From Instructors, 2) Course Participant Evaluation Results, 3) Degree of Professional Development Among Instructors, 4) Dollars Spent on Quality Assurance, 5) Maintenance Condition of the Motorcycles, 6) Quality of Instruction and Instructors, and 7) Quality of Instructor Updates.

Table 3

Criteria Named as Both Currently Used and Recommended for Ideal Use	
Capturing and Acting on Input From Instructors Course Participant Evaluation Results Degree of Professional Development Among Instructors	Dollars Spent on Quality Assurance Maintenance Condition of the Motorcycles Quality of Instruction and Instructors Quality of Instructor Updates

Table 4 presents the 17 criteria currently used for program evaluation within the framework of the CIPP Model. There were no criteria related to context evaluation, three criteria related to input evaluation, 10 criteria related to process evaluation, and four criteria related to product evaluation.

Table 4

CIPP Framework: Criteria Currently Used for Program Evaluation			
Context	Input	Process	Product
	Dollars Spent on Quality Assurance Maintenance Condition of Motorcycles Quality of Instruction and Instructors	Capturing and Acting On Input From Instructors Capturing and Documenting Unsolicited Responses From Students Degree of Assessing Individual Sites, Not State As a Whole Degree of Formal Documentation of Quality Degree of Professional Development Among Instructors Frequency of Instructor Updates Having a Skill Test Waiver Upon Course Completion Quality of Instructor Updates Requiring that Student Evaluations Be Completed Uniformity of Course Reporting	Accomplished Predetermined Goals Course Student Evaluation Results Numbers Trained Pass/Fail Rates in Courses

Table 5 presents the 30 criteria for ideal program evaluation in motorcycle safety programs within the framework of the CIPP Model. There were two criteria related to context evaluation, 11 criteria related to input evaluation, 13 criteria related to process evaluation, and four criteria related to product evaluation.

Table 5

CIPP Framework: Criteria For Ideal Program Evaluation			
Context	Input	Process	Product
Priority of Agency In Which Program Is Conducted Support From Superiors	Amount of Funding Clear Communication As the "Measuring Stick" for Program Dedication of People Involved Degree of Trust Within Program Dollars Spent on Quality Assurance Having Adequate Budget To Meet Expectations Having Adequate Number of Motorcycles Maintenance Condition of Motorcycles	Ability To Remedy/Remove Problems Within Program Capturing and Acting On Input From Instructors Degree of Emphasis On Service Function Degree of Professional Development Among Instructors Extent of Constant Learning Within Program Extent To Which Communication Is Open Extent To Which Program Participation Is Fun	Course Student Evaluation Results Increased Learning By Students Increased Skill By Students Training Itself Is Safe

Table 5 *(continued)*

CIPP Framework: Criteria for Ideal Program Evaluation			
Context	Input	Process	Product
	People Skills of Administrators and Instructors Quality of Instruction and Instructors Training Site Coverage	Extent to Which Quality Assurance Measures Are Implemented Extent to Which Quantity and Quality Are Raised Simultaneously Level of Communication Within Program Quality of Instructor Updates Quality of Interaction Between Instructors and Students Way Problems Are Addressed	

Further analysis of the data is presented in the context of the CIPP Model of program evaluation. A total of 18 primary questions were developed for the interviews with program administrators based on the CIPP Model of program evaluation. Five questions made up the context evaluation; four questions made up the input evaluation; five questions made up the process evaluation; and four questions made up the product evaluation.

Context Analysis. The five questions regarding context evaluation were: 1) How did your program come into being? 2) How is your program situated in the state system? 3) To what extent was a needs assessment conducted for your program? 4) What is the

administrative structure of your program? 5) How do you evaluate the administrative structure of your program?

Most of the programs were implemented because of legislative action. Most administrators were not in their current position when the legislation was passed, and could only speculate on the reasons for legislative interest. The administrators saw the establishment of a state motorcycle safety program as a trend across the county, and a few indicated that impetus came from organized motorcycle riding groups within the state.

The majority of programs are under the auspices of state departments affiliated with law enforcement and public safety. All program administrators indicated a strong and favorable relationship with their superiors. This ranged from being left alone, which was viewed as both favorable and unfavorable, to having a one-on-one relationship characterized by an open-door policy, which was viewed as favorable.

No program administrator was aware of a formal needs assessment being conducted prior to program implementation. It was assumed that the program was created to meet the needs of public demand for motorcycle training, and accident and injury reduction countermeasures.

The issue of administrative control was a theme regarding administrative structure. The range of control was from comprehensive state office oversight, to empowerment of sponsors and instructors. The degree of centralization varied. The administrative control was mostly associated with the student registration and instructor accountability processes.

There were no mechanisms for evaluating the administrative structure of the programs. The administrators agreed that little thought was given to how the state chose to situate the motorcycle safety program in state government. There was indications that a state office did not wish to have oversight regarding motorcycle safety programs. No administrator suggested an improved administrative structure that could enhance program outcomes. Most administrators liked the independence of program operations.

Input Analysis. The four questions regarding context evaluation were: 1) What are the primary objectives of your program? 2) Do you think your program is working up to its capabilities? 3) Is there any particular additions you would like to add to your program? 4) How do you evaluate the resources you use to achieve program goals?

The most common responses to the question regarding program objectives were that the program existed to train motorcyclists, expand the program, provide service, teach people how to ride and get people licensed, and improve motorcycle image and safety. A common theme was that the program should meet the minimum expectations of enabling legislation which was to promote safety and provide a service.

Administrators as a whole thought their programs were working up to perceived capabilities, and most related this to financial conditions. With more money, more riders

could be trained and service could be improved. All thought the service and instruction provided exceeded expectations and provided value for the investment.

Additional program features mentioned by administrators included acquisition more sites, procurement of more motorcycles, and availability of more certified instructors. These would provide the needed resources to train more of the public. Also, most administrators would like to provide better professional development opportunities for their instructors. This included personal development in human relations as well as technical assistance in delivering the curriculum.

There were no formal arrangements to evaluate the resources used in a program. The resources were finite, and administrators would make programmatic decisions to get the most out of a program with the limited resources provided. They had a concern about making improper administrative decisions insofar as complaints reaching their superiors.

Process Analysis. The five questions regarding process evaluation were: 1) Are there any design features that prevent your program from being more effective? 2) What percent of your time is devoted to program evaluation activities? 3) If you could change the way the program operates, what would be some of your recommendations? 4) How do you presently evaluate your program? 5) How do you evaluate the administrative aspects of your program?

The percent of time administrators spent thinking about program evaluation ranged from under five percent to over 70 percent. Several administrators would like to have outside evaluations conducted. Few felt qualified to formally evaluate program outcomes. Most administrators use student feedback to determine program effectiveness, and numbers trained as an indicator of program vitality. Administrators tended to evaluate independent program processes, such as instructor performance, instructor updates, and pass/fail rates, instead of overall program effectiveness. Key indicators were the frequency of student complaints and nature of instructor feedback. Evaluation at the student level was formalized through the use of course evaluation questionnaires, and evaluation from instructors was mostly informal through verbal communication. All administrators had an interest in improving the evaluation processes within their programs.

Product Analysis. The four questions regarding product evaluation were: 1) In what ways should your program be held accountable? 2) What outcomes do you measure? 3) What would you consider the ultimate measure of accountability for the effectiveness and efficiency of your program? 4) How do you evaluate the documentation you utilize to determine accountability?

Most administrators thought their programs should be held accountable to the degree the program outcomes met the letter of the law. This means to provide training for the public that desired access to the program. A few administrators mentioned meeting their superior's expectations as significant, while others commented that the ultimate goal was to reduce crashes, fatalities, and injuries.

The primary outcome that is measured is student satisfaction with the training course. All programs capture graduate information by using end-of-course evaluation procedures. Most administrators used these results to identify poor facilities, including motorcycles in need of repair, and to identify instructors that were not representing the program or not teaching the curriculum appropriately. Administrators would also gain feedback informally from the top, their supervisor, and from the bottom, their instructors. The primary quantifiable training data analyzed from year to year were numbers trained and the number of accidents/incidents during training.

Summary and Recommendations

The purpose of this qualitative research study was to identify criteria for program evaluation in motorcycle safety rider education and training programs from the perspective of program administrators. Two specific objectives were to identify current program evaluation practices by state administrators, and to identify program evaluation criteria for ideal practice.

Two hundred and five evaluation criteria were discovered from interviews with eight program administrators. Seventeen criteria were identified as presently being utilized by more than half of the administrators in program evaluation activities. A double-round Delphi procedure was used to identify criteria for ideal program evaluation practice. Thirty criteria were named by more than half of the respondents as having significant value for ideal program evaluation practice.

The following recommendations are offered: 1) There should be further exploration to refine criteria for program evaluation, 2) The relative importance of a criterion for ideal program evaluation should be determined, 3) Mechanisms and procedures to capture information relative to the criteria for ideal program evaluation should be explored, and 4) The development of a program evaluation model for motorcycle safety programs should be explored.

Appendix A

Initial List —Administrators' Ratings

Evaluation Criteria — Motorcycle Safety Programs

Criterion	**More Value**	**Of Value**	**Less Value**	**Cannot determine**	***Presently Used for Program Evaluation***
Ability to handle stand-by students	**ll**	**Lll**	**ll**		**ll**
Ability to handle walk-in students	**ll**	**Ll**	**lll**		**ll**
Ability to remedy/remove problems within program	**llll**	**Lll**			**ll**
Access to resources	**lll**	**Ll**			
Accident and fatality rates	**l**	**Lll**	**llll**		**llll**
Accomplishment of pre-determined goals	**lll**	**Lll**	**l**		**llll**
Addressing galvanizing questions (like helmet law)	**l**	**Lll**	**llll**		
Adequacy of finding emerging problems	**lll**	**Lll**	**l**		
Amount of advertising	**ll**	**Lll**	**lllll**		**l**
Amount of funding	**llllll**		**l**		**lll**
Amount of Instructor "bellyaching"	**l**	**Lll**	**ll**		**ll**
Amount of time administrator is free of procedural duties	**ll**	**llll**	**l**		
Amount of volunteer efforts within program	**l**	**llll**	**l**		
Assuring some riders discover motorcycling is not for them	**ll**	**Lll**	**ll**		
Availability and use of use of 402 funds	**lll**	**Lll**		**l**	**ll**
Balancing quantity of students and quality of program	**lll**	**llll**	**l**		**l**
Belief within program that motorcyclists must take care of their own		**llll**	**lll**	**l**	**l**
Capturing and acting on input from Instructors	**lllll**	**Ll**			**llll**
Capturing and documenting unsolicited responses from participants	**lll**	**llll**			**lllll**
Centralized administration	**ll**	**l1**	**ll**	**1**	**1**

Centralized registration system	**11**		**1111**	**1**	**1**
Clarity with which quality is defined	**11**	**11111**			**1**
Clear communication of the "measuring stick" for program	**11111**	**11**			**1**
Compliments to complaints ratio	**111**	**111**	**11**		
Conducting IPs only with CIs in state	**1**	**1111**	**1**	**1**	**1**
Coordination of individual training sites	**1**	**111**	**1**	**11**	**11**
Coordination of sites and course locations	**11**	**111**	**11**		**111**
Cost per rider		**111111**	**1**		**11**
Course student evaluation results	**1111**	**111**			**11111**
Courses produce learning in a Controlled environment	**111**	**11**	**1**	**1**	**11**
Decentralized administration		**11**	**11111**		
Dedication of the people involved	**1111**	**111**			**11**
Degree of assessing individual sites, not state as a whole	**111**	**1111**			**1111**
Degree of dealer involvement in program	**111**	**111**	**1**		**111**
Degree of distribution of information	**11**	**11**	**11**	**1**	**11**
Degree of emphasis on service function	**1111**	**11**		**1**	**111**
Degree of financial flexibility afforded administrators	**111**	**111**		**1**	**11**
Degree of follow up on riders that complete courses	**111**	**1**	**11**	**1**	**11**
Degree of formal documentation of quality	**11**	**1111**	**1**		**1111**
Degree of informal peer pressure among Instructors		**1111**	**11**	**1**	
Degree of Instructor control	**1**	**11111**	**1**		
Degree of meeting demand for training	**111**	**1111**			**11**
Degree of motorcycle enthusiast involvement		**11111**	**11**		
Degree of professional development among Instructors	**1111**	**111**			**1111**
Degree of public support	**11111**	**1**	**1**		**11**
Degree of reaching out beyond training courses	**1**	**1111**	**1**	**1**	**11**
Degree of scrutiny of program (less being better)	**1**	**111**	**11**	**1**	**1**

Degree of stakeholder feedback	**1111**	**1**	**1**	**1**	**1**
Degree of trust within program	**11111**	**11**			**1**
Degree of use of technology	**1**	**1111**	**1**	**1**	**1**
Degree to which administering agency understands program	**1111**	**111**			**1**
Degree to which best practices from other states are adopted	**11**	**111**	**11**		**11**
Degree to which change is accepted	**111**	**1111**			
Degree to which course graduates crash	**1**	**111**	**1**	**11**	**1**
Degree to which licensing and education dovetail (complement one another)	**1**	**111**	**11**	**1**	**1**
Degree to which lives are saved	**111**	**11**	**1**	**1**	**11**
Degree to which local sponsors have control	**1**	**11**	**111**	**1**	
Degree to which program environment is apolitical	**11**	**11**	**11**	**1**	
Degree to which there is "shared vision" within program	**11**	**11111**			**1**
Demand exceeds supply	**11**	**1111**	**1**		**1**
Direct and/or indirect involvement in lobbying efforts	**11**	**11**	**11**	**1**	
Diversity of sponsoring agencies	**11**	**1**	**11**	**11**	**1**
Documentation of year-to-year improvements exist	**111**	**1111**			**111**
Dollars spent on quality assurance	**1111**	**11**	**1**		**1111**
Ease of attracting ICs	**11**	**111111**			
Equipment has maximum utilization	**111**	**1111**			**11**
Evaluation occurs by advisory committee	**1**	**111**	**11**	**1**	**11**
Evaluation of only those items that can be controlled		**11**	**1111**	**1**	
Extent of constant learning within program	**11111**	**1**			**11**
Extent of Instructor recognition	**11**	**11111**			**11**
Extent of post-course follow up	**11**	**11**	**1**	**11**	**111**
Extent to which attitude of service is maintained	**111**	**1111**			**11**
Extent to which communication is open	**1111**	**111**			**1**
Extent to which courses are made available	**111**	**111**	**1**		**11**

Extent to which donations are solicited and collected	**1**	**111**	**111**		**1**
Extent to which evaluation is comprehensive, not single element	**111**	**111**	**1**		**11**
Extent to which evaluation is ongoing	**111**	**111**	**1**		**111**
Extent to which general public accepts/supports program	**11**	**11111**			**11**
Extent to which general public is educated about motorcycles	**1111**	**1**	**1**	**1**	**11**
Extent to which graduates think they are given a better head start	**111**	**1111**			**11**
Extent to which Instructors exceed minimum expectations	**111111**	**1**			**11**
Extent to which non-core courses are offered		**1111**	**11**	**1**	
Extent to which policymakers' expectations are exceeded	**111**	**111**		**1**	
Extent to which pre-testing/post-testing occurs	**11**	**11**	**1**	**11**	
Extent to which program participation is fun	**11111**	**1**	**1**		**111**
Extent to which program personnel/sponsors are motorcyclists	**11**	**111**	**11**		
Extent to which quality assurance measures are implemented	**111111**	**1**			**111**
Extent to which quantity and quality are raised simultaneously	**1111**	**111**			**11**
Extent to which self-evaluation is accomplished and enhanced throughout program	**11**	**11111**			**11**
Extent to which smaller markets are served	**111**	**111**		**1**	**11**
Extent to which sponsors are provided resources	**111**	**11**	**1**	**1**	**1**
Extent to which stakeholders can ask "Why" and "So What"	**111**	**1111**			**1**
Feedback from professional educators is solicited and documented	**11**	**11**	**1**	**11**	**1**
Focusing on process over focusing on results	**1**	**11111**	**1**		
Following national guidelines	**11111**	**11**			**11**
Formal annual reports are completed	**11**	**11111**			**111**
Formal evaluation of program administrator	**1**	**1111**	**1**	**1**	**1**
Frequency of Instructor observations	**111**	**1111**			**111**
Frequency of Instructor updates	**11**	**11111**			**11111**

Frequency of meeting with Instructors	**11**	**111**	**11**		**11**
Frequency of personnel turnover	**11**	**1111**	**1**		
Frequency of quality assurance visits	**1111**	**11**		**1**	**11**
Frequency of technical assistance visits	**11111**	**1**		**1**	**111**
Growth pattern	**111**	**1111**			**1**
Having a motorcycle traffic violator school	**11**		**111**	**11**	**1**
Having adequate budget to meet expectations	**111111**	**1**			**11**
Having adequate number of motorcycles	**111111**	**1**			**11**
Having adequate storage	**111**	**1111**			**1**
Having democratic, open communication	**111**	**11**	**11**		
Having incentives for course enrollment	**111**	**1111**			**111**
Having skill test waiver for course completion	**1111**	**11**		**1**	**1111**
Having standardized ranges		**11111**	**11**		
Having website to provide student/graduate feedback	**11**	**1111**	**1**		
Increase in licensed riders	**1111**	**11**	**1**		**1**
Increase in motorcycle licensure	**1111**	**11**	**1**		**1**
Increase in ratio of license holders/permit holders	**1111**	**11**	**1**		
Increased learning by participants	**1111111**				**11**
Increased skill of participants	**1111111**				**11**
Input of motorcycle groups	**111**	**1**	**111**		
Instructor longevity in program	**1**	**111111**			**1**
Instructors CPR/1st aid certified	**11**	**1111**	**1**		
Involvement of professional educators in program	**111**	**111**	**1**		**1**
Lack of complaints	**11**	**111**	**11**		
Level of communication within program	**111111**	**11**			**11**
Level of volunteering by Instructors	**1**	**11111**	**1**		**1**
Low number of no-shows	**11**	**111**	**11**		

Maintenance condition of the motorcycles	**1111**	**111**			**111**
Making it easy to obtain license for qualified riders	**111**	**11**	**1**	**1**	**11**
Membership in professional organizations	**111**	**11**	**1**	**1**	**1**
Minimal Instructor burnout	**111**	**1111**			
Mirroring of MSF guidelines/recommendations	**111**	**1111**			**111**
Network of communication among Instructors	**11**	**11111**			
Number of administrative meetings		**11**	**1111**	**1**	
Number of conferences attended by personnel		**1111**	**11**	**1**	
Number of full-time administrative personnel		**11111**	**11**		**1**
Number of full-time people in program		**11111**	**11**		
Number of motorcycle fatalities	**11**	**111**	**11**		**1**
Number of potential enrollees turned away	**11**	**1111**	**1**		
Number of site visits per year	**111**	**11**	**1**	**1**	**1**
Number/frequency of accidents in courses	**111**	**1111**			**111**
Numbers trained	**11**	**1111**	**1**		**11111**
On-line registration availability	**11**	**11**	**11**	**1**	
Participant cost for program	**111**	**111**	**1**		**11**
Pass/fail rate in courses	**11**	**11111**			**1111**
People skills of administrators and Instructors	**1111**	**111**			**11**
People's first impression	**1111**	**111**			**1**
Percent of course participants that get licenses	**111**	**111**		**1**	**1**
Percent of riders reached	**1**	**111**	**111**		
Philosophy and interest of program administrator	**11**	**11111**			
Priority of agency in which program is conducted	**11111**	**11**			
Professional development beyond classroom/range instruction	**111**	**1111**	**1**		**1**
Program structure is flat and non-hierarchical	**1**	**111**	**11**	**1**	
Providing re-tests without having entire course repeated	**1**	**1**	**1111**	**1**	**111**

Provisions for professional growth	**111**	**111**	**1**		**111**
Quality of fit in state traffic safety system	**1**	**11111**	**11**		**1**
Quality of instruction and Instructors	**11111**	**11**			**11111**
Quality of Instructor updates	**111111**	**1**			**1111**
Quality of interaction between Instructors and students	**111111**	**11**			**11**
Quality of original needs assessment		**11111**	**1**	**1**	**1**
Quality of promotional activities	**11**	**111**	**1**	**1**	**111**
Quality of record keeping	**111**	**1111**			**111**
Quality of screening Instructor candidates	**111**	**1**	**11**	**1**	**111**
Ratio of course participants to licensed riders	**1**	**11111**	**1**		**1**
Recognition of people involved in program	**1**	**11111**	**1**		**111**
Reduced emphasis on crashes and violations as an indicator	**1**	**111**	**111**		**11**
Reduces amount of illegal riding	**11**	**111**	**1**	**1**	**111**
Requiring that student evaluations be completed	**11**	**1111**		**1**	**11111**
Sense of community within program stakeholders	**11**	**1111**	**1**		
Separating "assessment of program" from "assessment of course participants"	**1**	**1111**	**1**		
Size of policies and procedures manual (larger being better)		**11**	**1111**		
Source of funding	**11**	**1111**			**1**
Stakeholder involvement in decision-making	**11**	**1111**	**1**		
State program compliance with the letter of the law	**1111**	**1**	**11**		**11**
Strength of local Instructors and coordinators	**111**	**1111**			
Sufficient number of motorcycles	**111**	**1111**			
Support from superiors	**11111**	**1**	**1**		
Tiered instructor system	**1**	**111**	**111**		
Time lines for Instructor follow up	**11**	**111**	**1**	**1**	
Training itself is safe	**1111**	**111**			**111**
Training site coverage in state	**1111**	**11**		**1**	**111**

Tuition is low	**11**	**11**	**111**		**1**
Uniformity of course reporting	**111**	**1111**			**1111**
Use of 3rd party testing	**1**	**111**	**11**	**1**	**11**
Use of advisory board	**1**	**1111**	**1**	**1**	**1**
Use of credit cards for tuition payment	**11**	**111**		**11**	**11**
Use of formal research	**11**	**11**	**1**	**11**	
Use of internet	**1**	**1111**	**11**		
Use of outside evaluators	**1**	**111**	**111**		**1**
Use of peer critiques	**1111**	**111**			**11**
Use of standardized assessments	**11**	**11111**			**1**
Use of student and graduate evaluations	**11**	**1111**	**1**		**111**
Use of website registration	**11**	**111**	**1**	**1**	**1**
Using background checks on Instructors	**111**	**11**		**11**	**11**
Videotaping is used for Instructor development	**11**	**11**	**1**	**11**	**11**
Visibility of program within the state	**11**	**11111**			**1**
Vision of administrators is growth-oriented versus controlling-and-monitoring-oriented	**11**	**11**	**111**		
Wait time for classes	**11**	**1111**	**1**		**1**
Wait time for classes less than 90 days	**1**	**1111**	**11**		
Way in which re-scheduling is accomplished	**11**	**1111**		**1**	**1**
Way in which re-tests are conducted	**1**	**111**	**1**	**11**	
Way problems are addressed	**1111**	**111**			**1**
Way staff is managed	**11**	**1111**	**1**		
Way students are dismissed	**111**	**1111**			**11**
Where program situated in state system	**111**	**1111**			**11**
Written and verbal testimonials are captured/documented	**1**	**1111**			**1**

Appendix B

Round 2 List —Administrators' Ratings

Evaluation Criteria — Motorcycle Safety Programs

Criterion	More Value	Of Value	Less Value	*Criterion not understood*
Ability to remedy/remove Problems within program	**11111**	**11**		
Accomplishment of pre-determined goals	**111**	**1111**		
Amount of funding	**111111**	**1**		
Capturing and acting on input from Instructors	**11111**	**11**		
Clear communication as the "measuring stick" for program	**1111**	**11**	**1**	
Course student evaluation results	**1111**	**11**	**1**	
Dedication of the people involved	**11111**	**11**		
Degree of emphasis on service function	**11111**	**1**	**1**	
Degree of professional development among Instructors	**11111**	**11**		
Degree of public support	**111**	**11**	**11**	
Degree of stakeholder feedback	**111**	**111**	**1**	
Degree of trust within program	**111111**	**1**		
Degree to which administering agency understands program	**111**	**111**	**1**	
Dollars spent on quality assurance	**1111**	**11**	**1**	
Extent of constant learning within program	**1111**	**1**	**11**	
Extent to which communication is open	**111111**	**1**		
Extent to which general public is educated about motorcycles	**111**	**111**	**1**	
Extent to which Instructors exceed minimum expectations	**11**	**11111**		
Extent to which program participation is fun	**1111**	**11**	**1**	
Extent to which quality assurance measures are implemented	**11111**	**11**		
Extent to which quantity and quality are raised simultaneously	**111111**		**1**	
Following national guidelines	**1**	**11111**	**1**	

Frequency of quality assurance visits	**1**	**1111**	**11**	
Frequency of technical assistance visits	**111**	**111**	**11**	
Having adequate budget to meet expectations	**11111**	**11**		
Having adequate number of motorcycles	**111111**	**1**		
Having skill tes t waiver for course completion	**11**	**111**	**11**	
Increase in licensed riders	**1**	**1111**	**11**	
Increase in motorcycle licensure	**1**	**111**	**111**	
Increase in ratio of license holders/permit holders	**1**	**111**	**111**	
Increased learning by participants	**111111**	**1**		
Increased skill of participants	**1111**	**111**		
Level of communication within program	**111111**	**1**		
Maintenance condition of motorcycles	**11111**	**1**	**1**	
People skills of administrators and Instructors	**11111**	**11**		
People's first impression	**111**	**1111**		
Priority of agency in which program is conducted	**1111**	**11**	**1**	
Quality of instruction and Instructors	**1111111**			
Quality of Instructor updates	**1111111**			
Quality of interaction between Instructors and students	**111111**	**1**		
State program compliance with the letter of the law	**111**	**1**	**111**	
Support from superiors	**1111**	**11**	**1**	
Training itself is safe	**11111**	**11**		
Training site coverage in state	**1111**	**1**	**11**	
Use of peer critiques	**111**	**111**	**1**	
Way problems are addressed	**1111**	**11**	**1**	

LIST OF REFERENCES

Baugher, Dan. (1981). Summary: Developing a successful measurement program. New directions for program evaluation, 11, 101-105.

Billheimer, John. (1997). The safety impact of driver education and training. The Chronicle, 45(2), 4-13.

Billheimer, J.W. (1996). California motorcyclist safety program. Program effectiveness: Accident evaluaton. California: California Highway Patrol.

Borich, Gary D. and Jemelka, Ron P. (1982). Programs and systems: An evaluation perspective. New York: Academic Press.

Boulmetis, John and Dutwin, Phyllis. (2000). The ABCs of evaluation: Timeless techniques for program and project managers. San Francisco: Jossey-Bass Publishers.

Buchanan, Lewis and Tarrants, William E. (1982). NHTSA technical report: Effectiveness and efficiency in motorcycle safety programs. Washington, D.C.: National Highway Traffic Safety Administration.

Cronbach, Lee and Associates. (1980). Toward reform of program evaluation: Aims, methods, and institutional arrangements. San Francisco: Jossey-Bass Publishers.

Cronbach, Lee F. (1963). Course improvement through evaluation. Teacher's college record, 64(8), 672-683.

Eisner, Elliot W. (1985). The educational imagination: On the design and evaluation of school programs. New York: MacMillan Publishing Company.

Elliot, W. (1983). Educational connoisseurship and criticism: Their form and functions in educational evaluation (pp. 335-247). In Evaluation models: Viewpoints on educational and human services evaluation. Boston: Kluwer-Nijhoff Publishing.

ERS Standards Committee. (1982). Evaluation research society standards for program evaluation. New directions for program evaluation, (15), 7-19.

Fetterman, D.M. (1995). In response to Dr. Daniel Stufflebeam=s: Empowerment evaluation, objectivist evaluation, an evaluation standards: Where the future of evaluation should not go and where it needs to go. Evaluation Practice, 16(2), 179-199.

Fetterman, D.M., Kaftarian, S. and Wandersman, A. (1996). Empowerment evaluation: Knowledge and tools for self-assessment and accountability. Thousand Oaks: Sage Publications.

Fetterman, D.M. (1996). Empowerment evaluation: An introduction to theory and practice, in Fetterman, D.M., Kaftarian, S., and Wandersman, A. (1995). Empowerment evaluation: Knowledge and tools for self-assessment and accountability. Thousand Oaks, CA: Sage Publications.

Guba, E.G. (1969). The failure of educational evaluation. Educational technology, May, 29-38.

Guba, E.G. and Lincoln, Yvonna, S. (1989). Fourth generation evaluation. Newbury Park: Sage Publications.

Guba, Egon G. and Lincoln, Ynonna, S. (1983). Epistemological and methodological biases of naturalistic inquiry. In Madaus, George F., Scriven, Michael S. and Stufflebeam, Daniel L. (Eds.). Evaluation models: Viewpoints on educational and human service evaluation (311-333). Boston: Kluwer-Nijhoff Publishing.

Guba, E. and Lincoln, Y.S. (1981). Effective evaluation. San Francisco: Jossey-Bass Publishers.

Higham, S. (1980). Education is the key. Driver, 14(4), p. 8-9.

House, Ernest R. (1983). Assumptions underlying evaluation models. In Madaus, George F., Scriven, Michael S. and Stufflebeam, Daniel L. (Eds.). Evaluation models: Viewpoints on educational and human service evaluation (pp. 45-64). Boston: Kluwer-Nijhoff Publishing.

Jonah, B.A., Dawson, N.E., and Bragg, B.W.E. (1982). Are formally trained motorcyclists safer? Accident analysis and prevention, 14(4), 247-256.

Kirkpatrick, Donald L. (1994). Evaluating training programs. San Francisco: Berrett-Koehler Publishers.

Knowles, Malcolm S. (1980). The modern practice of adult education. Englewood Cliffs: Cambridge Adult Education.

Kosecoff, Jacqueline and Fink, Arlene. (1982). Evaluation basics: a practitioner's manual. Beverly Hills: Sage Publications.

Lonero, Lawrence P. and Clinton, Kathryn M. (1998). Changing road user behavior. Toronto: PDE Publications.

Madaus, George F., Scriven, Michael S., and Stufflebeam, Daniel L. (1983). Evaluation models: Viewpoints on educational and human service evaluation. Boston: Kluwer-Nijhoff Publishing.

Mayhew, D.R. and Simpson, H.M. (1996). Effectiveness and role of driver education and training in a graduated licensing system. Ottawa, Canada: The Traffic Injury Research Foundation.

McDavid, J.C., Lohrmann, B.A., and Lohrmann, G. (1989). Does motorcycle training reduce accidents? Journal of safety research, 20(2), 61-72.

McKnight, A.J. (1987). Evaluation of the Pennsylvania motorcycle safety program, Final report. Indiana, Pennsylvania: University of Pennsylvania.

McLaughlin, John A. & Associates (1998). New directions for program evaluation, (39), 1-6.

McPherson, K. (1989). Motorcycle licensing: A look at where we are. Journal of traffic safety education. 34(3), 6-8.

Merriam, Sharan B. (1998). Qualitative research and case study applications in education. San Francisco: Jossey-Bass Publishers.

Meyers, William R. (1981). The evaluation enterprise. San Francisco: Jossey-Bass Publishers.

Mortimer, R.G. (1984). Evaluation of the motorcycle rider course. Accident analysis and prevention, 16(1), 63-72.

Mortimer, R.G. (1988). A further evaluation of the motorcycle rider course. Journal of safety research, 19(4), 187-196.

Motorcycle Industry Council. (1992). Motorcycle statistical annual. Irvine, CA: Motorcycle Industry Council.

Mowbray, Carol T. (1998). Getting the system to respond to evaluation findings. New directions for program evaluation, 62(4), 47-58.

National Association of State Motorcycle Safety Administrators. (1998). Motorcycle safety assessment: Guidelines and objectives.

National Highway Traffic Safety Administration. (1999). The art of appropriated evaluation. Washington, D.C.: Department of Transportation.

National Safety Council. (1998). Accident Facts. Itasca, IL: National Safety Council.

Osga, G.A. (1980). An investigation of the riding experiences of MSF rider course participants in South Dakota. Report HFS-80-2, South Dakota: University of South Dakota.

Pfeiffer, John. (1968). New look at education: Systems analysis in our schools and colleges. New York City: The Odyssey Press.

Philippi, Jorie W. (1996). Basic workplace skills. The ASTD training & development handbook: A guide to human resource development (pp. 819-843), Fourth Edition. New York: McGraw-Hill.

Rothe, J. P. and Cooper, P.J. (1987). Motorcyclists: Image and reality. British Columbia: Insurance Corporation of British Columbia.

Satten, R.S. (1980). Analysis and evaluation of the motorcycle rider course in thirteen Illinois counties. In Proceedings of the international motorcycle safety conference, pp. 145-193. Linthicum, MD: Motorcycle Safety Foundation.

Scriven, M. (1967). The methodology of evaluation. AERA monograph series on curriculum evaluation, No. 1: Perspective on curriculum evaluation. Chicago: Rand McNally.

Shadish, William R., Newman, Dianna L., Scheirer, Mary Ann, and Wye, Christopher (Eds.). (1995). Guiding principles for evaluators. New directions for program evaluation, (66).

Simpson, H.M. and Mayhew, D.R. (1990). The promotion of motorcycle safety: Training, education, and awareness. Health education research, 5(2), 257-264.

Smith, N. (Ed.). (1981). Metaphors for evaluation. Beverly Hills, Calif.: Sage Publications, Inc.

Stake, Robert E. (1967). The countenance of educational evaluation. Teacher's college record, 68(8), 523-540.

Stake, Robert E. (1983). Program evaluation, particularly responsive evaluation. In Madaus, George F., Scriven, Michael S. and Stufflebeam, Daniel F. (Eds.). Evaluation models: Viewpoints on educational and human service evaluation (287-310). Boston: Kluwer-Nijhoff Publishers.

Steele, Sara M. (1991). The evaluation of adult and continuing education. Handbook of adult and continuing education. San Francisco: Jossey-Bass Publishers.

Stufflebeam, D.L. (1995). Empowerment evaluation, objectivist evaluation, an evaluation standards: Where the future of evaluation should go and where it needs to go. Evaluation practice, 15(3), 321-338.

Stufflebeam, Daniel L. (1972). The relevance of the CIPP evaluation model for educational accountability. SRIS quarterly, 3-6.

Stufflebeam, Daniel L. (1983). The CIPP model for program evaluation. In Madaus, George F., Scriven, Michael S. and Stufflebeam, Daniel L. (Eds.). Evaluation models: Viewpoints on educational and human service evaluation (pp. 117-141). Boston: Kluwer-Nijhoff Publishers.

Stufflebeam, Daniel, et. al. (1971). Educational evaluation and decision making. Itasca, Ill: Peacock.

Stufflebeam, Daniel L. and Webster, William J. (1983). An analysis of alternative approaches to evaluation. In Madaus,, George F., Scriven, Michael S., and Stufflebeam, Daniel L. (Eds.). Evaluation models: Viewpoints on educational and human service evaluation (pp.23-43). Boston: Kluwer-Nijhoff Publishers.

Thompson, Nancy J. and McClintock, Helen O. (1998). Demonstrating your program's worth: A primer on evaluation for programs to prevent unintentional injury. Atlanta: National Center for Injury Prevention and Control.

Tyler, R.W. (1949). Basic principles of curriculum and instruction. Chicago: University of Chicago Press.

Weiss, Carol H. (1997). Theory-based evaluation: Past, present, and future. New directions for evaluation, (76), 41-55.

Winn, G.L. and McPherson, K. (1990). Developing motorcycle program evaluation criteria. Proceedings of the 1990 international motorcycle safety conference, Vol 1: The human element (pp. 6-97, 6-106). Irvine, CA.: Motorcycle Safety Foundation.