Criteria for program evaluation in motorcycle safety rider education and training programs

Raymond Joseph Ochs

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss

Recommended Citation
To the Graduate Council:

I am submitting herewith a dissertation written by Raymond Joseph Ochs entitled "Criteria for program evaluation in motorcycle safety rider education and training programs." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Education, with a major in Education.

John M. Peters, Major Professor

We have read this dissertation and recommend its acceptance:

E. Grady Bogue, Ralph G. Brockett, Walter A. Cameron

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)
To the Graduate Council:

I am submitting herewith a dissertation written by Raymond Joseph Ochs entitled: “Criteria for Program Evaluation in Motorcycle Safety Rider Education and Training Programs.” I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree Doctor of Education, with a major in Education.

John M. Peters, Major Professor

We have read this dissertation and recommend its acceptance:

E. Grady Bogue
Ralph G. Brockett
Walter A. Cameron

Accepted for the Council:

Interim Vice Provost and Dean of The Graduate School
CRITERIA FOR PROGRAM EVALUATION IN MOTORCYCLE SAFETY
RIDER EDUCATION AND TRAINING PROGRAMS

A Dissertation
Presented for the
Doctor of Education
Degree
The University of Tennessee

Raymond Joseph Ochs

August, 2001
ABSTRACT

The purpose of this study was to identify program evaluation criteria for motorcycle safety rider education and training programs from the perspective of state program administrators. The two research questions were: "What are the currently-used criteria for program evaluation in motorcycle safety rider education and training programs?" and "What are ideal criteria for program evaluation in motorcycle safety rider education and training programs?" Eight administrators were interviewed using questions developed within the framework of Stufflebeam’s Context-Input-Process-Product (CIPP) model of program evaluation. A total of 205 criteria were initially identified from an analysis of the interview results data, and administrators named 17 of these criteria as currently used for program evaluation. Through the use of a double-round Delphi technique, administrators rated 30 criteria as being ideal for program evaluation.

Comparison of ideal criteria to criteria in current use revealed differences in qualitative features and scope of administrator concerns, such as learning outcomes and instructor performance. Ideal criteria may prove useful in the development of evaluation systems for motorcycle safety rider education and training programs, but further study of these criteria is warranted. Needed is a better understanding of barriers to implementation of ideal criteria and whether the criteria identified in other regions of the country parallel those identified in this study.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Statement of the Problem</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Purpose of the Study</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Significance of the Study</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Limitations</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Organization of the Dissertation</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>Review of Related Literature</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Part One. History and Development of Program Evaluation</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Part Two. Evaluation of Motorcycle Safety Programs</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Part Three. Models of Program Evaluation</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>Research Methodology</td>
<td>46</td>
</tr>
<tr>
<td>4</td>
<td>Findings</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Part One. Currently-used Criteria for Program Evaluation</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Part Two. Ideal Criteria for Program Evaluation</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Part Three. Comparing Currently-used and Ideal Criteria</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>Summary and Conclusions</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>Summary</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>Discussion</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>Conclusions and Implications</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>Appendixes</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>Appendix A: Evaluation Criteria: Round One with Frequency of Responses</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td>Appendix B: Evaluation Criteria Round Two with Frequency of Responses</td>
<td>141</td>
</tr>
<tr>
<td></td>
<td>Appendix C: Administrator Letter of Invitation</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>Appendix D: Administrator Second Letter</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>Appendix E: Form B: Human Subjects Form</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>Appendix F: Informed Consent Statement</td>
<td>154</td>
</tr>
<tr>
<td></td>
<td>Vita</td>
<td>157</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>CIPP Framework: Currently-used Criteria for Program Evaluation</td>
<td>55</td>
</tr>
<tr>
<td>Table 2</td>
<td>CIPP Framework: Ideal Criteria for Program Evaluation</td>
<td>74</td>
</tr>
<tr>
<td>Table 3</td>
<td>Criteria Both Currently Used and Ideal for Program Evaluation</td>
<td>100</td>
</tr>
<tr>
<td>Table 4</td>
<td>Criteria Currently Used But Not Ideal for Program Evaluation</td>
<td>102</td>
</tr>
<tr>
<td>Table 5</td>
<td>Ideal Criteria Not Currently Used for Program Evaluation</td>
<td>103</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

During the 1960's, there was an alarming increase in the frequency and severity of highway accidents and collisions. The number of motor vehicle deaths rose from 38,137 in 1960 to 54,633 in 1970 (National Safety Council). In response to this rising toll, then President Lyndon Johnson signed into law a comprehensive bill, commonly referred to as the Highway Safety Act of 1966. This legislation established eighteen standards, including recommendations for safety programs that state agencies could adopt at their discretion. Motorcycle Safety was one of the program standards, and its purpose was "to assure that motorcycles, motorcycle operators and their passengers meet standards which contribute to safe operation and protection from injuries" (National Highway Safety Bureau, 1966).

Also during this period, there was a startling increase in motorcycle sales and registrations, and an increase in motorcycle accidents, fatalities, and injuries. In 1960 motorcycle registrations totaled 575,000, increasing to 2,815,000 by 1970 (Motorcycle Industry Council, 1992). Fatalities during this period rose from approximately 700 to over 2,000 during this same time period (United States Government Accounting Office, 1991). This rise in motorcycle accidents lead to the formation of the Motorcycle Safety Foundation (MSF), which was funded by the five leading motorcycle distributors in the United States (Harley-Davidson, Honda, Kawasaki, Suzuki, Yamaha). Founded in 1973, the MSF named as its mission the reduction of motorcycle accidents and injuries through promotion of safe and responsible motorcycle operation. The cornerstone of its efforts
was the development of a motorcycle rider education and training program for new motorcyclists. The MSF introduced a learn-to-ride curriculum, the Beginning Rider Course (BRC), in 1974. It was replaced in 1976 with the Motorcycle Rider Course (MRC), and an updated version of this curriculum was published in 1985 called the Motorcycle RiderCourse: Riding and Street Skills (MRC:RSS).

Programs were offered around the country through local sponsors, such as community colleges, universities, and motorcycle dealerships. But course offerings were sporadic, and there was little in the way of organizational structure at state or local levels. The United States Department of Transportation's National Highway Traffic Safety Administration, in partnership with the MSF, urged states to pass legislation that would establish funding for motorcycle safety programs. In 1979, Rhode Island became the first state to implement a state-funded motorcycle safety program. To date 46 states have legislation to operate and maintain formal motorcycle safety rider education and training programs.

Statement of the Problem

No comprehensive evaluation plan exists specific to motorcycle safety rider education and training programs. The efficacy of such programs was assumed, but literature identified a need for purposeful program evaluation activities. "A review of available literature revealed no existing survey of state motorcycle programs regarding evaluation. One survey which addressed needs of state programs was conducted by the State Motorcycle Safety Coordinator's Council (1988). That document recounted an investigation of many aspects of individual state programs but did not address evaluation"
(Winn and McPherson, 1990). The authors recommend that state administrators take a fresh look at the utility of program evaluation while being mindful of recent legislative attempts to re-appropriate user fees for purposes other than motorcycle safety.

Purpose of the Study

The purpose of this study was to identify criteria for program evaluation in motorcycle safety rider education and training programs. The study had two specific objectives: 1) to identify currently-used criteria for program evaluation, and 2) to discover ideal criteria for program evaluation. The primary research questions were: 1) What are the currently-used criteria for program evaluation in motorcycle safety rider education and training programs? and 2) What are ideal criteria for program evaluation in motorcycle safety rider education and training programs?

Significance of the Study

As programs continue to develop, and as quality, efficiency, and effectiveness are challenged by stakeholders and funding sources, the development of criteria for program evaluation can lead to the development of a model of program evaluation for such programs.

This study's findings should contribute to the traffic safety literature and to the motorcycle safety education and training profession. Safety education and training programs are developed and implemented for the purpose of reducing the frequency and severity of traffic crashes by influencing human factors. Educational programming, in concert with engineering and enforcement techniques, can have an impact on accident
rates. But quantitative research studies do not demonstrate a significant impact in terms of reductions in traffic accidents and violations. Alternative means of program evaluation need to be developed. Lonero and Clinton (1998) note "...expert feelings on the benefits gained from safety education seem to range from nervous to strongly skeptical. Sound evaluations and demonstrated effectiveness are relatively rare for any type of road safety education" (p. 41). The identification of ideal criteria for program evaluation can serve as a basis for program administrators and others to more effectively evaluate motorcycle safety rider education and training program operations and results.

Limitations

Limitations of the study included the following:

1) The findings of this study were limited to the states that make up the southeast region of the United States Department of Transportation's National Highway Traffic Safety Administration.

2) The criteria for program evaluation were based upon the perceptions of program administrators.

3) The interpretation of the data was from the perspective of the researcher.

Organization of the Dissertation

Chapter Two of the dissertation contains a literature review in three parts. Part One provides a review of the history and development of educational program evaluation; Part Two provides a summary of scientific studies specific to evaluation of motorcycle safety education and training programs; Part Three provides a review of several
educational program evaluation models. Chapter Three consists of a description of the research methodology for this study. Chapter Four presents the findings that include the currently-used criteria and the criteria identified as ideal for program evaluation. Chapter Five contains the summary, discussion, conclusions, and the implications of this research.
CHAPTER 2
REVIEW OF 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 from the following 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) noted 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.
A growing emphasis regarding program evaluation was expanding its uses. Stake (1983) noted: "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. These principles focused on procedures 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 were 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. 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 later added to these 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).

Borich and Jemelka (1982) offered another perspective regarding the development of program evaluation. They viewed the primary purpose of evaluation as
traditionally providing decision makers with information about the effectiveness of some program, product, or procedure, and argue that differentiating definitions hamper the development of the discipline. After reviewing how various fields and how terminology is used to conduct program evaluation, they present a distilled definition. “Despite disparity in the terms used across fields, all of these definitions refer to some act of supplying information appropriate to some premeditated question, the answer to which is believed to have practical utility for subsequent decision making” (p. 2).

This perspective was challenged by Fetterman (1995). He responds to comments about objectivist-type evaluations and states: “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).

According to Borich and Jemelka (1982), a milestone in the history and development of program evaluation came in 1965 with the passage of the Elementary and Secondary Education Act by the United States congress, which provided funds for educational materials and research. Evaluations were tied to funding in an effort to make programs accountable. For the first time educators were required to devote time and resources to evaluating their own effort. But evaluations failed to prove what empirically was observed, and few formal evaluations demonstrated significant results.

In an effort to add structure to the program evaluation profession, the Evaluation Research Society (ERS), established in 1976, was established to serve the professional needs of the growing number of practitioners engaged in program evaluation. “The ERS
believes that even though evaluators have different titles, work in different areas, and come from different backgrounds, they have common concerns and interests; and that evaluation theory and practice will benefit from interdisciplinary sharing. These beliefs have led to a search for standards to guide program evaluation practice and to focus attention on issues facing the emerging profession" (ERS Standards Committee, 1982, p. 8).

The ERS Standards Committee established categories for the range of potential evaluation services. They stated that while some people tend to think of program evaluation as a one-shot effort to determine the overall worth of a program, in fact this is only one of several general categories of evaluation. The six categories are front-end analysis (preinstallation, context, feasibility analysis), evaluability assessment (to determine if other kinds of evaluation should be initiated, formative evaluation (developmental, process) to test or appraise the processes of an ongoing program in order to make modifications and improvements, impact evaluation (summative, outcome, effectiveness), program monitoring to acknowledge that evaluation is ongoing as opposed to a one-shot event, and evaluation of evaluation (secondary evaluation, metaevaluation, evaluation audit, may include utilization evaluation).

The ERS Standards Committee organized the 55 standards into six sections. These were Formulation and Negotiation, Structure and Design, Data Collection and Preparation, Data Analysis and Interpretation, Communication and Disclosure, and Utilization.

Borich and Jemelka (1982) provided three general orientations for educational evaluation. These are the educational psychology models, the educational decision
models, and the educational science models. The educational psychology models are characterized by identification of behavioral objectives, administering measurement tools, and comparing results with the objectives. The educational decision models are characterized by a process of ascertaining the areas of concern, discovering information about these areas, and presenting useful information to decision makers. The educational science model, using the classical methodology of experimental control over variable, studied the causal relationships of the inputs of a program, the program’s operation or treatment using the inputs, and outcomes of the program. Because the dimensions of program evaluation are varied and multidimensional, the complexity in clearly identifying incisive results can be overwhelming. Program evaluation had a propensity for pragmatic decision making more so than theoretical value judgments. “Rather the evolution of this state of affairs has resulted from the belief that evaluation’s central role is to respond to the pragmatic demands of the marketplace and to the needs of those in most immediate control of the program” (Borich and Jemelka, 1982, p. 24).

Presently the field of evaluation is diffuse. Educational programming is often deeply contextual and, if evaluated, requires idiosyncratic evaluation methodologies. Guba and Lincoln (1989) viewed evaluation in terms of three generations, and offered 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 described 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 the simple measurement of 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) offered another perspective regarding the motivations for
conducting program evaluation, particularly in the realm of social programs intended to
ameliorate social problems. He asked: “Is there a need for program evaluation?” He believed the changes in recent times reflect political developments rather than advances in evaluation methodology. He suggested 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) provided recommendations appropriate for developing a successful measurement program. He identified 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.

Cronbach and associates (1980) argued for a comprehensive transformation of program evaluation. They urged that program evaluation not have as a purpose to eliminate the fallibility of authority or bolster 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 provided 95 theses as principal points to spur debate regarding the fundamental issues surrounding program evaluation.

Utilization in program evaluation was a galvanizing topic for the newly formed American Evaluation Association. In their first national conference in 1986, the primary topic of concern was utilization. “Utilization has many faces. At times uses are easy to identify—that is, indicators are explicit, allowing the evaluator and program staff to logically link evaluation recommendations to program decisions. Or to put it another way, changes in programs that can be attributed to evaluation information can be observed. Resources are increased or decreased. Interventions are modified or terminated. New programs replace old programs” (McLaughlin and Associates, 1988, p. 1).

Mowbray (1988) suggested that reasons for nonutilization could be grouped into two categories: blame the evaluator or blame the decision maker; and additionally, there must be a differentiation between instrumental uses and conceptual uses of evaluation findings. “While instrumental uses of evaluation results may be relatively rare, conceptual uses can have a long-term and significant impact on policy decisions, how problems are defined, and in setting the overall policy directions that are compatible with accumulated knowledge” (p. 47-48).

In 1992, the American Evaluation Association (AEA) established a task force to develop guiding principles for evaluators. This differed from the work of the ERS standards committee. Instead of developing a competing list of operational principles, a more general listing of conceptual principles was developed. Five overarching principles were identified that included 23 specific principles. The five overarching principles were
systematic inquiry (evaluators conduct systematic, data-based inquiries about whatever is being evaluated), competence (evaluators provide competent performance to stakeholders), integrity/honesty (evaluators ensure the honesty and integrity of the entire evaluation process), respect for people (evaluators respect the security dignity, and self-worth of the respondents program participants, clients, and other stakeholders with whom they interact), and responsibilities for general and public welfare (evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare (Shadish and Associates, 1995).

Program evaluation literature runs the gamut from theoretic perspectives to how-to-do manuals. A recent line of inquiry in program evaluation is theory-based evaluation. This perspective promotes the idea that evaluation can provide a much greater and broader contribution if it is so designed to study the mediating mechanisms between process and outcome. Weiss (1997) suggests that theory-based evaluation can pursue two different strands. "One path is to build more detailed program theories, so that evaluations can trace the micro-steps of process all along the pathways that lead to program effects...The other path that theory-based evaluation can pursue is to limit the theory to the one or two central assumptions embedded in each program" (p. 51-52).

Two recent publications provide program evaluation ideas for traffic safety practitioners, which would include motorcycle safety program 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 was 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 suggested 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 was 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 were solid research designs with random assignment to experimental and control groups and sophisticated statistical analysis. But exception was 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).

The practical implementation of program evaluation activities in the traffic safety community has, to some degree, excluded evaluation experts. Although not referring directly to traffic safety programming, Boulmetis and Dutwin (2000) provided 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 were 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 program evaluation in 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 in response: “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. The findings do not show a difference in crash rates.

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. It was found that when controlling for age and years licensed, course graduates did not have a lower accident rate.
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) reported no statistically significant differences in crash rates as a function of different training and licensing systems. McPherson (1989) noted 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 license applicants were matched on the basis of age, sex, and prior driving records. There were no significant differences were found between the mileage-adjusted collision rates of trained and untrained riders.

Rothe and Cooper (1987) commented on evaluations of motorcycle education programs. They note: "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 suggest also attending to the process of how the course is presented and how learning is experienced by the participant.

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 noted 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
assumption 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 pointed 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) noted: “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). The results of this study continue to be scrutinized.

Simpson and Mayhew (1996) included a commentary in their review of several studies related to motorcycle safety programs. They stated: “In summary, several studies have failed to provide definitive conclusions about the effectiveness of rider education and training in reducing crashes” (p. 34).

Fassnacht (1997) believed rider education programs provide a meaningful experience, but acknowledges the lack of proof. He stated: “Yet as strongly as I feel about it I know, with frustration, that they all have something in common: none has yet to produce the equivocal proof that it works” (p. 30).

The review of research regarding motorcycle safety rider education and training programs indicated that formal evaluation studies have not produced a clear conclusion regarding the effectiveness of these programs. While there was some evidence of knowledge and skill improvement, a significant reduction in the incidence of motorcycle crashes could not be shown.
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. Borich and Jemelka (1982) note: “Models are efforts to simplify and render phenomena understandable, and sometimes they lose fidelity in the process of simplification” (p. 177-178).

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 is the professional field in general” (p. 69).

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 was directly tied to the concept of curriculum objectives. He stated: “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 a 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

Stake Countenance Model

The Countenance Model involved 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

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 an 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 encouraged the examination of relationships to the process. This involved 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. He suggested that evaluators not consider program objectives when conducting an evaluation, and 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 was offered by Meyers (1981) as he stated: "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) connected 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 was 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) pointed out the primary advantage and disadvantage. They stated: "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 did 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 were 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).

Stake's Responsive model differs from his Countenance model in two distinct ways. One is that it is less complicated, which was a criticism of his earlier model. The other is that Stake's Responsive model is more closely associated with a constructivist viewpoint; that is, this type of evaluation is emergent and designed to capture unintended effects.

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) named 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 referred 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 process consisted 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) stated: “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 acknowledged 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) offered 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 considering contexts where training and education may be synonymous, 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 saw 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.

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 defined evaluation as the process of delineating, obtaining, and providing useful information for judging decision alternatives (Stufflebeam, 1971, p. xxv). “Fundamentally, the use of the CIPP Model was 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). Four kinds of decisions were specified: 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. The model divided evaluation into four distinct strategies—Context evaluation, Input evaluation, Process evaluation, and Product evaluation, thus the acronym CIPP (Borich and Jemelka, 1982, p. 10).

**Context** evaluation provided information about the needs, problems, and opportunities in order to identify objectives. “Its purpose is to provide a rationale for determination of objectives” (Stufflebeam and others, 1971, p. 218). It defined 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 provided 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 provided information about the strengths and weaknesses of a strategy during implementation so that 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 provided 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).

One of the criticisms of the CIPP model was that it ignored summative evaluation due to its emphasis on fostering improvement. Although the CIPP Evaluation Model was originally developed to provide timely information in a systematic way for decision making, it can function as a retroactive purpose of providing information for accountability (Stufflebeam, 1972, p. 3). Characteristics of the model are: (1) It treats evaluation as a systematic, continuing process; (2) The evaluation process separates the types of questions and information needed for decision making; and (3) That decision
making is intended to improve programs. “It meshes well with traditional organizational training evaluation and with the value-added training evaluation model” (Philippi, 1996, p. 838).

The CIPP model was chosen for this study based on the above characteristics and because it provides a system’s view of education and training. It provides a framework that is sufficiently structured to assure coherence in analyzing the data results, and it is sufficiently flexible to allow probes into the operational aspects of motorcycle safety rider education and training program activities. The CIPP model is consistent with the best features of many of the other models. For example, Madaus, Scriven, and Stufflebeam (1983, p. 124) note that a full implementation of CIPP requires the evaluator to ask such questions as the following:

1. What needs were addressed, how pervasive and important were they, and to what extent were the project’s objectives reflective of assessed needs (addressed by context evaluation)?

2. What procedural and budgeting plan was adopted to address the needs, what alternatives were considered, why was it chosen over them, and to what extent was it a reasonable, potentially successful, and cost effective response to the assessed needs (addressed by input information)?

3. To what extent was the project plan implemented, and how and for what reasons did it have to be modified (addressed by process information)?

4. What results – positive and negative as well as intended and unintended – were observed, how did the various stakeholders judge the worth and merit of the outcomes, and to what extent were the needs of the target population met (product information)?
The CIPP model accommodates a range of choices in planning and implementing program evaluation activities. Stufflebeam (1983) acknowledged this when he stated: “Evaluations...need to carefully plan their evaluation activities so they can carry them out efficiently and with an acceptable amount of rigor and convince their clients that they know what they are doing. On the other hand, they need to approach the design of evaluation studies flexibly and provide for periodically reviewing and otherwise modifying the design so that the evaluation remains responsive to the needs of the audiences” (p. 136). Thus the CIPP model provides evaluators with a flexible mechanism to adapt to the changing demands of program evaluation.

Galvin (1983) surveyed the training community to measure its attitude regarding the CIPP model and the Kirkpatrick model in management training. He discovered that the CIPP model is preferred by more training specialists, primarily if used to evaluate broad approaches to curriculum development. “The CIPP model, which was developed for professional education use, can make a valuable contribution to corporate training. The CIPP model is practical, effective at improving management training, cost efficient and comprehensive. It provides a balanced view of evaluation, and it is useful for short seminars and long-range management development programs” (p. 57).

In an update of the CIPP model, Stufflebeam (1983) provided 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 (p. 140).

Brookfield (1986) pointed 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).

Summary

This chapter provided a review of the literature related to educational program evaluation. It traced the history of program evaluation, and showed the various stages of development as program evaluation evolved and as new models were introduced. Early models were shown to be empirically-based forms of measurement with an emphasis on pre-post evaluations, when modern models are qualitative in form. The complicated nature of program evaluation was evident in a review of the range of evaluation models.

A review evaluation research specific to motorcycle safety programs revealed that no study had definitely demonstrated the safety value of motorcycle safety rider
education and training programs. The motorcycle safety community is perplexed by the lack of proof of these safety training initiatives.

A review of ten program evaluation models included both quantitative and qualitative perspectives. The CIPP model was discussed in terms of its particular advantages as a framework for the present study.
CHAPTER 3
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 are Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, and Tennessee. This region was chosen as a convenience sample, and it represented states in proximity to the University of Tennessee, Knoxville, and to the domicile of the researcher, Kentucky. An administrator in each of the sample states was identified utilizing resources from the Motorcycle Safety Foundation and the 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 in this study. 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. Each of the eight program administrators that were invited agreed to this process. The open-ended interviews with
The program administrators lasted between one and two hours, and were audio-taped and transcribed. The tapes were destroyed after transcription.

The same primary questions were asked of all administrators. Probing questions varied depending on the answers and the need for clarification. The primary questions were piloted by interviewing two program administrators from outside the sample. This was to assure that the primary questions were appropriately understood. The 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 has characteristics appropriate for addressing evaluation criteria for program evaluation in motorcycle safety rider education and training programs. The CIPP model is designed for educational programming and divides evaluation into four distinct types: context evaluation, input evaluation, process evaluation, and product evaluation. Stufflebeam (1971) explains:

- **Context evaluation** serves planning decisions to determine objectives; **input** evaluation serves structuring decisions to determine project designs; **process** evaluation serves implementing decisions to control project operations; and **product** evaluation serves recycling decisions to judge and react to project attainments. To monitor a system and thereby provide information on needed changes, context evaluation is mainly general and systematic. The other three types of evaluation are specific and *ad hoc*; they come into play only after a planning decision has been reached to effect some sort of system change, and
specific evaluation designs for each vary according to the setting for the change (p. 218).

Context evaluation is the most basic type. It defines the program environment, describes actual and desired conditions, identifies opportunities, and diagnoses the problems that prevent needs from being met. According to Stufflebeam (1971): “Context evaluation has many distinguishing characteristics. It is macro analytic—it sets the boundaries of the system to be evaluated and then describes and analyzes it” (p. 219). It serves to specify operational context and to identify underlying needs. It is designed to continuously assess needs, problems, and opportunities within a decision maker's domain.

Input evaluation provides information to determine how to utilize resources to meet program goals. Stufflebeam (1971) notes: “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” (pp. 222-223). The process is microanalytic, and it provides the capability to assess needs, opportunities, and problems. It is used to identify and assess system capabilities. It focuses on assessing alternative means for achieving specified ends.

Process evaluation provides information related to program decisions. It looks at interpersonal relationships among staff and clients, communication channels, logistics, shared vision, program intent, and adequacy of the resources, physical facilities, staff, and time requirements. According to Stufflebeam (1971), process evaluation "provides project decision makers not only with information needed for anticipating and overcoming procedural difficulties but also with a record of process information for
interpreting project attainments” (p. 232). He emphasizes the congruency of processes and intended means, suggesting adjustments and refinements that may be called for by actual experience.

The purpose of product evaluation (Stufflebeam, 1971) is “to measure and interpret attainments not only at the end of a project cycle, but as often as necessary during the project term. 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” (p. 232). This evaluation is concerned with comparing actual and intended effects, as well as unintended effects.

The CIPP model was appropriate for this study because it provided a framework to enable the researcher to identify currently-used criteria for program evaluation and discover ideal criteria for program evaluation. No comprehensive evaluation plan currently exists for motorcycle safety rider education and training programs, and the criteria discovered can form a basis from which to develop a model of program evaluation. Borich and Jemelka (1982) note that the “CIPP model allows the measurement of program outcomes both during and after program completion. At each stage outcomes are compared to stated objectives and differences between expected and actual results are reported to decision makers” (p. 10). Also in support of using the CIPP model for this study, Guba and Lincoln (1981) note: “Evaluation within the CIPP model is thus a process for delineating, obtaining, and applying descriptive and judgmental information concerning some object’s merit as revealed by its goals, structure, process,
and product. In addition, it is a process undertaken for some useful purpose such as
decision making or accountability” (p.15-16).

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: “What are the primary objectives of your
program?” “Do you think your program is working up to its capabilities?” “Is 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: “Are there any design features that prevent
your program from being more effective?” “What percent of your time is devoted to
program evaluation activities?” “If you could change the way the program operates, what
would be some of your recommendations?” “How do you presently evaluate your
program?” “How do you evaluate the routine processes of your program?”

Product evaluation questions were: “In what ways should your program be held
accountable?” “What outcomes do you measure?” “What would you consider the
ultimate measure of accountability for the effectiveness and efficiency of your program?”
“How do you evaluate the documentation you utilize to determine accountability?”

From responses to these primary questions and probing questions, the researcher
identified program evaluation criteria by using a constant comparative method. This
method is described in Merriam (1998): “Categories and subcategories (or properties) are most commonly constructed through the constant comparative method of data analysis...at the heart of this method is the continuous comparison of incidents, respondents’ remarks, and so on, with each other. Units of data—bits of information—are literally sorted into groupings that have something in common” (p. 179).

Answers from program administrators were analyzed as data were collected, and again as the transcripts were reviewed. From the analysis of data, a total of 205 evaluation criteria were identified (Appendix A).

Evaluation criteria were identified by analyzing the transcripts after the interviews were completed. The CIPP model was valuable in establishing the primary interview questions as well as the probing questions that lead to the discovery of 205 criteria. Any criterion that related to program evaluation that was mentioned or referred to was included in the initial list of 205 criteria. For instance one administrator, in responding to a question regarding how he would like to evaluate his program, stated the following:

Well, I think we need to do a couple of things. I think we could have better feedback from the students. I think that the method we have is not one of the best, and I’m not sure what the best is, but I think you really need to get that feedback. And I think sometimes a better way is to do some follow up after the student has been out of the class for a while. Typically, I think most people when they go through the course, even if they’ve been riding for a number of years, find the course interesting, challenging, and rewarding; unless you get some problem, maybe it’s extremely cold or hot...problem with a motorcycle or problem with an instructor, something of that nature. But typically my experience has been that
they really enjoy the course. So at the end of the course if you ask them to fill out a questionnaire, they’re going to be real positive about what’s going on in general. And I think that there should be some assessment of the student at the period maybe three or six months or twelve months after that. Did they get their license? Did they get a motorcycle and were they riding? Did they still ride on a regular basis? Just some of that. What from the course helped you the most? What helped you the least? Get some of those kinds of evaluations. I think the other area we need to work on besides the students is, I think, we need to work with the instructors a little bit, especially with technical assistance, and provide additional evaluation of what they’re doing and feedback. One of the things we did that I found to be very effective, I think instructors like it after some initial introduction, was to do some videotaping of the instructors. Most of them found it very helpful, especially if you’d do some kind of running commentary with them in the classroom or on the range.

Evaluation criteria gleaned from the above statement included the following: Course student evaluation results, degree of follow up on riders that complete courses, frequency of technical assistance visits, increase in licensed riders, quality of instruction and instructors, and videotaping is used for instructor development. Comments and concerns voiced by administrators that were related to program evaluation were considered as potential evaluation criteria.

To answer the first research question “What are the currently-used criteria for program evaluation in motorcycle safety rider education and training programs?” an alphabetized list of the 205 evaluation criteria was mailed to each administrator in the
sample. They were asked to identify any criterion that was currently used for program
evaluation. A criterion was selected as a “currently-used criterion” if at least one-half of
the administrators selected a particular criterion. A total of 17 were identified as
currently-used criteria for program evaluation. A criterion was identified as currently-
used if more than one-half of the respondents indicated that a particular criterion was
used for program evaluation.

To answer the second research question “What are ideal criteria for program
evaluation in motorcycle safety rider education and training programs?” a two-round
Delphi technique was employed. For the first round, each administrator was mailed the
alphabetized list of the 205 criteria and asked to rate each one as to its value in program
evaluation (more value/of value/less value) based on accuracy, clarity, and utility. If a
criterion was rated as having “more value” by more than one-half of the administrators,
that criterion was considered for the second round. Forty-six criteria were rated as
having “more value” in the first round. A second-round Delphi technique was employed
to further refine the 46 criteria. An alphabetized list of the 46 criteria was mailed to each
administrator again asking them to rate the value of each criterion. A criterion was
named as an ideal criterion if it was rated as having “more value” than the others by at
least one-half of the administrators. A total of 30 criteria were identified as ideal criteria
for program evaluation.
CHAPTER 4
FINDINGS

This chapter is divided into three parts. Part One provides the 17 currently-used criteria for program evaluation. Part Two provides the 30 criteria identified as ideal for program evaluation. Part Three provides the seven criteria that were both currently-used and ideal, as well as the criteria that were currently used but not ideal and the ideal criteria not currently used.

Part One. Currently-Used Criteria in Program Evaluation

To identify currently-used criteria, a listing of the 205 criteria were mailed to each administrator, asking each of them to note which of the evaluation criteria were currently utilized for their particular program. A criterion was identified as currently-used if more than half of the respondents indicated that a particular criterion was used for program evaluation. Seventeen criteria were identified as currently-used criteria.

Table 1 presents, within the framework of the CIPP model, the 17 currently-used criteria for program evaluation. A description follows each criterion.

There were no criteria related to context evaluation, three criteria related to input evaluation, ten criteria related to process evaluation, and four criteria related to product evaluation.
<table>
<thead>
<tr>
<th>Context</th>
<th>Input</th>
<th>Process</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Dollars Spent on Quality Assurance</td>
<td>Capturing and Acting On Input From Instructors</td>
<td>Accomplishment of Predetermined Goals</td>
</tr>
<tr>
<td></td>
<td>Maintenance Condition of Motorcycles</td>
<td>Capturing and Documenting Unsolicited Responses From Students</td>
<td>Course Student Evaluation Results</td>
</tr>
<tr>
<td></td>
<td>Quality of Instruction and Instructors</td>
<td>Degree of Assessing Individual Sites, Not State As a Whole</td>
<td>Numbers Trained</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Degree of Formal Documentation of Quality</td>
<td>Pass/Fail Rates in Courses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Degree of Professional Development Among Instructors</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Frequency of Instructor Updates</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Having a Skill Test Waiver Upon Course Completion</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quality of Instructor Updates</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Requiring That Student Evaluations Be Completed</td>
<td></td>
</tr>
</tbody>
</table>
Context Evaluation: Currently-used Criteria

There were no currently-used criteria for context evaluation in these motorcycle safety programs. This is likely due to the fact that programs were mandated by law.

Input Evaluation: Currently-used Criteria

There were three currently-used criteria for input evaluation in these motorcycle safety programs.

1. Dollars Spent on Quality Assurance. This criterion refers to an emphasis on quality versus quantity. Although administrators expressed that the numbers trained is a primary objective of a statewide program, the level of quality experienced by course participants is paramount. Quality assurance means to have the resources and the instructional competency to satisfy customers. There is a tension that exists between quality assurance and numbers trained, but attention to the amount of money spent on quality assurance assumed a level of quantity that meant service is provided at a minimal level.

Adam: *Oh, it* (referring to weaknesses resulting from quality assurance visits) *was mostly in the instructional quality. While the cat's away, the mice will play. If nobody's watching or if nothing happens about something.*
Dan: *I would throw it (additional funds) at instructors, in their development. Doing a lot of stuff that MSF is doing right now. I'd love to see them expand what they're doing in terms of professional development, teaching techniques. Get them the tools, some things they can use to improve our program. Because, I'm convinced if the program works, it's got to be working through the instructor.*

Edward: *I think the other area we need to work on besides the students is I think we need to work with the instructors a little bit, especially with technical assistance, and provide additional evaluation of what they're doing and feedback. One of the things we did in _____ that I found to be very effective, I think instructors like it after some initial introduction, was to do some videotaping of the instructors.*

2. **Maintenance Condition of Motorcycles.** This criterion refers to the operational condition of the motorcycles as well as their cosmetics. Motorcycle breakdown results in less program participation. Motorcycles that are damaged or dirty reflect on program image.

Adam: *Oh yeah, one of my real problems right now is I believe the maintenance and the upkeep of motorcycles is a strong indicator, as I tell my site coordinators: We have only one opportunity to make a first impression, and if those motorcycles don't make the right impression when the students arrive at the range on Saturday morning, they start off with the wrong impression.*
Carl: ...we do all the maintenance on the motorcycles. Unless it's just like a motor job or something like that...we have the typical dings, dents, broken mirrors, turn signals, levers and so forth; and, we evaluate those because we do two things: one, if something happens on the bike, the instructor is required to fill out our incident report form. He's also to inspect the motorcycle and if there's anything wrong with the motorcycle, he fills out a form and sends it back.

3. Quality of Instruction and Instructors. This criterion refers to dynamics of instructor, participant, and curriculum. Administrators assume the curriculum is sound because it was developed and approved by the Motorcycle Safety Foundation. The way the curriculum is presented from a technical standpoint, and the characteristics of the instructor/participant experience are fundamental to the quality of the program. In fact, this was one of two criteria that received a unanimous “high value” rating.

Adam commented on what he looks for in instructor candidates: You know, desire and attitude; attitude is what I'm really looking for. I also look at their previous...how long they've been doing whatever it is they're doing. I usually come up with something silly, you know, tell me about...if they're a fireman or whatever, tell me about how you roll up a fire hose. Just to see if they can make a short presentation; same thing that we were taught years ago in the military.

Ben: I think it's the dedication and quality of our group of instructors...we all seem to have the same common interest of providing a safety message and
offer some training assistance to individuals and see them go and prosper and be happy and enjoy their motorcycling.

Carl: I think the strongest component of our program would have to be our instructors. I'm not saying that just to blow smoke, but I'm around these people, I know them on a personal level... Our second strongest point, I think, is that we do a good job in the classes and we get so many referrals for word-of-mouth.

Dan: We brought some of these new teaching techniques in. Some of them (instructors) are resisting them like they are the plague. And, man, they're great. And it's getting out of the mode to try it. I think some different eyes and some different areas of expertise need to look at what we're doing.

Edward: Well we do have a real need for instructors...and that they have the skills to be good instructors later on. Some of our sponsors, like I said, are really good about recruiting...So I'd say that's an issue that's going to be an ongoing issue for a little while until we can work on that and kind of get our numbers up.

Frank, in referring to program strengths: They (instructors) understand quality. They understand in this state we operate strictly by MSF guidelines, and they've got a tried-and-true program; and, my demand is that we adhere to that.

Hal: The strongest components or elements in this particular program lie with the instructor ranks. Realizing that the stronger instructor base...what I mean by a stronger instructor base is not numbers, numbers are good, but it's
their vast wealth of knowledge and determination to help meet the needs of the students.

Process Evaluation: Currently-used Criteria

There were ten currently-used criteria for process evaluation in these motorcycle safety programs.

1. Capturing and Acting on Input from Instructors. This criterion refers to gaining feedback from the instructors in the program. Smaller programs more frequently communicate informally with instructors via telephone or email. All administrators welcomed input they received from instructors, and recognized the significant value of instructor perceptions. One of the most valuable activities within program operation is the annual instructor updates that programs conduct.

Dan: Instructors talk to me in a different way. I think we've given them some of the best equipment, some of the nicest ranges. We've standardized all the range painting. We paint all of them the same way. They have an excellent crew that does all that. They complain to me...what I look at is their longevity and their growth. And when I see some folks stay with us a long time and see they're still growing, then I know we're doing something for that person.

Frank: And we do an annual survey at our instructor update where we survey them for their input; you know, how do they rate overall how the program is being operated; how do they rate the particular update as far as how it was run; was it effective; did you take something home out of this or was it several
hours that you just had to be there so you could stay on the state-approved instructor list; what would you like to see change within the program, as far as how it is run; what you think would improve something or what do you think is a waste of time. We get input back from instructors...I'd say that most always the good ideas come from instructors. I mean, that's the backbone of any good organization, is the soldiers so to speak. They're the ones out there in the field all the time that are actually doing this thing.

Hal: As a matter of fact, several of our sites, we're getting testimony from instructors and probably thinking out loud about it now, we probably need to document those testimonies from instructors. Keep them on file of the number of students that came back and kept saying I was in this situation and here's how I responded, and here's how I reacted.

2. Capturing and Documenting Unsolicited Responses Students. This criterion refers to the testimonials received by administrators and/or instructors from students who completed the training program. These testimonials were both written and verbal.

Adam: Favorably. We're right up there now. We're starting to get some...we got just recently...we got some real positive feedback from unsolicited responses from participants. When I showed them to the Captain, he said in all the years he's been there, he's never seen anything, any reports.

Carl: Well, we try to have a presence at...we do the evaluation in class. We get letters back from people, by and large, overwhelmingly positive. We try to keep a presence in the area at motorcycle events and get feedback that
way...word of mouth is the thing. We get people that call in and say 'My friend took it."

Dan: Well, like I told you, the feedback I get from students. Not only reading their evaluations but I field all the calls, the complaints. They all come to us. And all the good stuff comes to us too. I keep a file on complimentary letters on instructors and the program. Feedback from the instructors I certainly get. If we've got problems with a sponsor, I hear them. Any problem they have in conducting the class. We make a big point in making sure they know that we know. Otherwise, we can't help them out there as they do their job. So we get a lot of feedback both from instructional staff and students. Get feedback from the public, in terms of public information campaigns as well. I hear it all.

Greg: I have had people come to me and say 'Look, this happened to me a couple of weeks ago and I remember what you taught me. And all I say is that I'm glad I had the training because I probably wouldn't be here if I hadn't.' I've had instructors come back to me and tell me, you know, some short story like that. They're sitting in the restaurant eating with their wife or whatever. And someone will be staring at them. And they turn around to figure out why this guy is staring at them and a few minutes later come over and introduce himself and say we are here today because of your training. We want to just thank you and let you know. So, yeah, I think it's really doing...it's hard to evaluate that and give a true reality value to how much we are worth. But I really think the rider education is educating people and saving lives.
Some of the students will email me. I'm getting more and more email from the students after class wanting to compliment the program. Wanting to compliment the instructors. I used to get a lot of letters, I'd say 10-15 letters every year, maybe up to 20 letters written by students. And now I'm getting a lot of email.

3. **Degree of Assessing Individual Sites, Not Entire State.** This criterion refers to how an administrator maintains a sponsor for the rider education program. Some sponsors are more efficient and effective than others, and an administrator does not want a poor sponsor to reflect negatively on the whole state program.

   Carl: *One of the recurring problems is just...you know, I sit and cringe a lot of times every year when somebody calls and it's somebody from one of our sites. We have a tough time finding and keeping sites. Again, it's not because they don't like what we're doing. But things change. We lost our classroom site at ___ because the state said that they were going to have to move all of their portable classrooms. Wasn't any fault of theirs. They, in fact, helped us find another site. We have moved probably four times in ____ with our site. Because we were at the police academy and they decided to make a parking lot. We left there and went with the city and the city decided they needed to make a parking lot. We left there and went with the city and the city decided they needed the parking space and built a softball complex.*

4. **Degree of Formal Documentation of Quality.** This criterion refers to keeping records that transcends the mere recording of numbers trained, but to create a formal process that focuses on the quality of the program and its offerings.
Dan: I'd call it informal at this point. Like I told you, we tried some very formal evaluation and some very in-depth type of things to look at. And now it's more, I think, informal. People just want to see us train people. There's a cry for more classes. And more, more, more. And the quantity pushes on. And we've got to answer the quantity issues but we can't do that by sacrificing the quality of the program.

Hal: I'd like to see us form, I don't like the word formal... but I would like to see us have an advisory committee that is represented by key individuals such as one or two people from the instructor ranks and one or two people from _____ and then from the _____ that we're operating with, from groups, clubs, organizations. Yet to be real informal; to say here's the concepts.

5. Degree of Professional Development Among Instructors. This criterion refers to amount of formal and informal continuing education that instructors accrue after they graduate from the motorcycle safety instructor preparation course. Administrators think that their instructors are the corner stone of the program, and that as instructors continue to become better, the program itself will improve. The primary interaction within program operation is the dynamic that occurs between the course participants and instructors.

Dan: Well, from my perspective and looking at everything we do, I'd still have to go to the instructors, and I'd like to have some evaluative tool or component built into the program where they could grow, where they could maintain their interest and be damn good at what they do.
6. **Frequency of Instructor Updates.** This criterion refers to how current instructors are interfacing with program administration, particularly in curriculum matters. This may take the form of an annual conference or be a one-on-one observational visit.

    Adam: *We originally up until about two years ago, were doing three separate updates, one for the east, one for the central, and one for the west—starting two years ago, and initially I pushed to have a single, unified update. We’ve had two of them so far.*

    Edward: *We require instructors to attend at least one update every two years. And in the past, they’ve done those with groups of people at different times throughout the year, and kind of rotated that every two years, where they bring in half a dozen or so and do an update that then wait three or four or five months and then bring in another dozen or so and do an update, and do that.*

    Frank: *We do site quality audits. And we do an annual survey at our instructor update where we survey them for their input. You know, how do they rate overall and how the program is being operated, how do they rate the particular update as far as how it was run—was it effective. Did you take something home out of this or was it several hours that you just had to be there so you could stay on the state-approved instructor list. What would you like to see changed in the program, as far as how it is run.*

    Greg: *They’re looked at and I try to update them annually. As we go through, you were talking about evaluation earlier, I’ve made some changes to*
reduce the stress on the students and reduce instructor stress. I’ve developed say, some improvements.

Hal: And then I’ll also do what we call an observation visit, instead of the technical assistance visits which tends to apply to real formal evaluation. The observation visit is to come by...it’s a little less threatening. So that is being done.

7. Having a Skill Test Waiver Upon Course Completion. This criterion refers to the mechanism whereby a course graduate bypasses the regular licensing test procedure, and simply receives a motorcycle license by showing the department of motor vehicles a course completion card.

Greg: And then we’ve added two hours of practicing the state skills test. The state has three different standards to go through the serpentine. And that’s the biggest drawback of passing the state test...and only one person has told me that they hadn’t attempted the test because they were still intimidated by it. Everybody else I’ve talked to has said they passed it the first time around with minimum points deducted. So we are helping people get licensed. And actually by offering this course, it’s helping them get licensed, despite themselves. We’re educating them and I feel making them a safer rider.

There’s no correlation in the state at this particular time as far as DMV waiver or third party testing. That is in the process—one of our goals. Hopefully by the end of the year the instructors will be classified as safety officers, and the Level II evaluation (end of course skill test for the basic course) will count for a waiver of the actual motorcycle test with the DMV.
I expect my numbers and demand will double once there's third party testing, because there's so many people in the state operating on learner's permits that cannot pass the test.

8. Quality of Instructor Updates. This refers to the conduct of annual instructor updates. Administrators put significant emphasis on the development of their instructors. Other than quality assurance visits to observe instruction, administrators have little opportunity to provide formal professional growth among their instructors. An instructor update is the primary mechanism for an administrator to promote personal growth with direct interfacing. This criterion was one of two criteria that received a unanimous "high value" rating.

Carl: We do one-day updates, and they will consist of...we require all of our instructors to stay CPR and first aid certified; and, it may be that we do a half day of...our last one was a half day of CPR certification for those that didn't have it. And the last part was MRC (Motorcycle RiderCourse) and ERC (Experienced RiderCourse) problems and administration. And it may be that we come back and the next one we do will be simply an update of the MRC to make sure we're all on the same page and so forth.

Greg: We do evaluate and provide the instructors during the annual update...we show the national, pretty much what we get from MSF, the national standards, and we actually break down and can give every instructor their data on their students' performance as far as the knowledge test, which questions they
missed, an exactly where they took points off of the skill evaluation; and, then we
do it by state.

9. Requiring That Student Evaluations Be Completed. This criterion refers to
having course participants complete an end-of-course evaluation form at the end of the
course.

   Ben: I did have a form that both use, the instructor and the participant
use. It is a kind of a general, yes/no or comment. If they want to they can write
out a comment. There is no signature required, nobody has to sign it. It's given
to them at the completion of the course...and there's been some real good input
out of that. By listening to the people that are interested in the program or
listening to the participants—and as a result, we've improved.

10. Uniformity of Course Reporting. This criterion refers to having each site that
conducts a course utilize the same student evaluation form. The uniformity facilitates
gaining a true picture of overall student evaluation of a course, and provides a mechanism
for a comparative analysis by administrators.

   Carl: We don't have a bunch of numbers and not do anything with them.
   We look at the end. We track what county's they're from. We track all around.
   We look at the average student. What's an average student? We track their age
   and all that.

   Edward: Well, we have all of our sponsors complete a student report form,
and we do collect all the registrations...we're hoping that that will help us
identify, even geographic areas, maybe where we're not getting people coming to
the course, and how best to determine locations.
Frank: I do things, you know; when you start talking about numbers and
looking at that sort of thing. Every class that we run, I get a class report that
comes in. I get student evaluations of the instructors and of the program. I look
at every one of those when they come in. My administrative assistant puts all this
information into a database and periodically throws out some charts or
spreadsheets. She gives me something to look at.

Product Evaluation: Currently-used Criteria

There were four currently-used criteria for product evaluation in these motorcycle
safety programs.

1. Accomplishment of Pre-determined Goals. This criterion refers to meeting the
regulatory requirements as set forth by a governing agency through statutory laws and/or
administrative regulations.

Adam: The easy answer is that it has to comply with state law. But that’s
easy to comply with as it is written. Everything about this program is...not
everything...about the basic structure is the _____ law annotated. It tells what an
instructor is, tells what a chief instructor is, what a site is. It tells what all this is.
I mean, that’s the easy part, complying with that part of it.

Dan: I don’t know...I could tell you we’re spending money to provide
training, to improve licensing. We’re not just loading the money up and shooting
it up in the air. But, I don’t know. I feel like we’re doing what the intent of the
law is, anyway, or the letter of the law...I always look at our instructors. They
seem to be the key to things. If we can keep people ten, fifteen years in this
program teaching and being active and growing, I feel like we're doing what we're supposed to do; if we keep getting the feedback from students that we get that it's one of the best training programs that they've ever been in. And we've had some trainers and educators actually call our office and write our office telling us that, I feel like we're doing what we're supposed to be doing. I don't have anything else to hang my hat on.

Edward: Well, I think we're directly accountable for the quality of the program. I think that's clearly our responsibility that we administer the program and that's what the law requires us to do. So the quality of the program, that directly relates to us and it's roughly what we're trying to do.

Greg: It should be accountable fiscally and to show quality and constant improvement. Looking for ways of improving--that we don't have a lot of complaints against the program. All in all, the students are happy with it.

Hal: We've already met the objectives of the state. My personal objectives or the philosophy of the team approach that we have, I'd say we're better than 50 percent—we're on the other side of reaching our goals.

2. Course Student Evaluation Results. This criterion refers to end-of-course evaluation forms that are completed at the end of each class. Administrators expressed that reviewing student evaluations is the primary outcome that is evaluated in demonstrating program effectiveness. The information acquired from these evaluations ensures that participants are satisfied with their experience, that the instructor performed adequately, and the facilities and equipment were adequate to meet program and student needs.
Ben: I did have a form that both use, the instructor and the participant uses. It is kind of general yes/no comment. If they want to they can write out a comment. There are no signatures required, nobody has to sign it. It’s given to them at the completion of the course. After they have graduated, after they have gotten their completion cards. We’re not holding bribe over them, say you don’t pass if you don’t fill it out right. And we honestly want to have their input, both of them. And there’s been some real good input out of that. By listening to the people that are interested in the program or listening to the participants. And the result, we’ve improved.

Edward: Right, every student should receive a survey form. And they get an envelope too and are requested to take the form home and fill it out, and then drop it in the mail. And that’s been marginally successful. A lot of students don’t fill it in and typically, they’ll go one-way or the other. Either they’re really bad, like this course was just horrendous for whatever reason, or they’re really good.

Greg: It allows me to...it’s good information. It gives me feedback on the instructors and gives me feedback on the facilities; and, it we have used it to change facilities.

3. Numbers Trained. This criterion refers to the number of graduates from the rider education and training program.

Carl: We measure the number of students that we teach.

Dan: I guess our effectiveness file, as it stands right now, is our statistics file—numbers trained, numbers of crashes. We just look at raw data. It doesn’t say anything. It just says...it’s all growing. It’s getting bigger.

71
Greg: We report by items on the contract. What's contracted there. One is PR visits, instructors trained, the number of sites being taught, the number of classes. The number of students enrolled.

4. Pass/Fail Rates in Courses. This criterion refers to the number of participants that pass the course, and the number of participants that are unsuccessful.

Ben: Well, basically as you know, they have to pass a written knowledge test and a skill evaluation. I look at the scores. And from time to time, if I'm not involved in the class and get a chance to drop in on the instructor teaching the class, I like to see the classroom presentation. Just a few minutes of it or the whole thing. Same way on the range. When they're teaching and particularly the evaluations on the range—to see if they're grading what they should be grading, not just looking over mistakes and not discounting points for them.

Frank: ...we'll look at pass/fail rates and drop rates within a class. We look at the state average...we can look that up on an individual instructor basis. For example, I had one site I was looking at recently and their drop or failure rate was about three times what the state average was. Of course, that threw up a red flag.

Greg: And we actually break down and can give every instructor their data on their students' performance as far as the knowledge test, which questions they missed. And exactly where they took points off of the skill evaluation. And then we do the state, so they can compare their standard for evaluating Level II evaluations with the state average. And the best we can figure, the state is running pretty close to the national average, as far as the points deducted.
Part Two. Ideal Criteria for Program Evaluation

To identify the ideal criteria for program evaluation, a double-round Delphi was employed, asking administrators to rate each criterion as having “more value,” “of value,” or “less value” than the others. The first round contained the 205 criteria gleaned from the interviews. If a criterion was named as having “more value” by at least half of the respondents in that round, it was considered for the second round. The first-round Delphi yielded 46 criteria that had “more value.”

A second-round Delphi yielded 30 criteria that were rated as having “more value.” Appendix B is the list of the 46 criteria and tallies from the second-round Delphi.

Of the thirty criteria identified as ideal, there were two criteria related to context evaluation, eleven criteria related to input evaluation, thirteen criteria related to process evaluation, and four criteria related to product evaluation.

Table 2 presents within the framework of the CIPP model, the 30 criteria identified as ideal for program evaluation by the administrators. A description follows each criterion, except for those that were previously described as a currently-used criterion.

Context Evaluation: Ideal Criteria

There were two ideal criteria for context evaluation in these motorcycle safety programs.

1. Priority of Agency in Which Program Is Conducted. This criterion refers to how the motorcycle safety program is situated within the state system. Administrators were sensitive to the importance placed on their programs by their superiors and by the
Table 2

CIPP Framework: Ideal Criteria For Program Evaluation

<table>
<thead>
<tr>
<th>Context</th>
<th>Input</th>
<th>Process</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority of Agency In Which Program Is Conducted</td>
<td>Amount of Funding</td>
<td>Ability To Remedy/Remove Problems Within Program</td>
<td>Course Student Evaluation Results</td>
</tr>
<tr>
<td>Support From Superiors</td>
<td>Clear Communication As the “Measuring Stick” for Program</td>
<td>Capturing and Acting On Input From Instructors</td>
<td>Increased Learning By Students</td>
</tr>
<tr>
<td></td>
<td>Dedication of People Involved</td>
<td>Degree of Emphasis On Service Function</td>
<td>Increased Skill By Students</td>
</tr>
<tr>
<td></td>
<td>Degree of Trust Within Program</td>
<td>Degree of Professional Development Among Instructors</td>
<td>Training Itself Is Safe</td>
</tr>
<tr>
<td></td>
<td>Dollars Spent On Quality Assurance</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Having Adequate Budget To Meet Expectations</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Having Adequate Number of Motorcycles</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maintenance Condition of Motorcycles</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

74
Table 2 (Continued)

<table>
<thead>
<tr>
<th>People Skills of Administrators and Instructors</th>
<th>Extent To Which Quality Assurance Measures Are Implemented</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of Instruction and Instructors</td>
<td>Extent To Which Quality and Quantity Are Raised Simultaneously</td>
</tr>
<tr>
<td>Training Coverage in State</td>
<td>Level of Communication Within Program</td>
</tr>
<tr>
<td></td>
<td>Quality of Instructor Updates</td>
</tr>
<tr>
<td></td>
<td>Quality of Interaction Between Instructors and Students</td>
</tr>
<tr>
<td></td>
<td>Way Problems Are Addressed</td>
</tr>
</tbody>
</table>

administration of the state agency. They think that the extent to which their program is a priority reflects the emphasis on quality measures as well as allocation of funds.

Carl: To tell you the truth, us being small as we are, I'm not sure that within the state government that many people know about us. But there are a number of people that assist our program whenever they can. People from the state safety coordinator committee have been very good to the program.
Dan: ...I've always felt like we weren't a priority with this agency. Not that this agency hasn't helped this program. This agency has made sure that we've had funding and they've worked to help us in the program; but, when you're a civilian inside of what they call a paramilitary organization, you're just not a priority.

Edward commented on his program's position: I think we're in our niche where we're at. And I don't see that changing much. I mean, I don't have other outside groups, rather stakeholders, or other people wanting to do more or do less. I don't see anybody really making any opposition to the program.

Hal: But the changes I would make have less to do with administrative structure as it does with administrative attitudes toward motorcycling in general. Not the structure itself. The way we're set up the 'pecking order' that we have established or the hierarchy as it breaks down has worked effectively only because the people who are in those positions, you know, work well together.

2. Support From Superiors. This criterion refers to importance placed on both the motorcycle safety program itself and the amount of trust and credibility given administrators by their immediate superiors. Administrators expressed some concern regarding program evaluation that reflected upon them personally, when in fact the operational aspects and results of the program may not be in their control.

Adam: It's basically turned over to me, but he (the supervisor) has been very supportive in making personal appearance at events. He and I have gone
and talked to the Commissioner of Safety together, so he is a supportive individual.

Dan: Yeah, we’ve made a lot of changes over the years. I call it zigzagging; the zigzag approach. This didn’t work too well so let’s try something else. We’re always evaluating, trying to figure out ways...we started the program like a lot of programs are set up now with state coordinators and site sponsors and it just didn’t work for us. We couldn’t keep somebody in those key locations long enough. Or the school itself or whoever we were working with, they’re goals and stuff changed and we weren’t really part of their program, per se. I always felt like we were sort of like a stepchild. And then we’d be struggling to find another location, another site coordinator. And that doesn’t go on as much now that we’ve kind of gone to this centralized approach. And this has worked better for us in terms of managing all the different locations across the state.

Frank: ...my immediate supervisor leaves me alone. It’s my program. He lets me run it as long as everything seems to be running efficiently. He’s there if I need him. I just report in occasionally and say ‘Hey, we’re going to be gone a few days.’ If anybody is looking for us we’re going to be here or be there. We sit down occasionally and find the direction we’d like to go in. He goes over those things with us. He kicks in some ideas. He pretty much lets us run it. It’s not a lot of red tape that we have to go through or a lot of politics that we have to deal with.
Greg: Very little supervision, very little contact. As long as everything seems to be running smooth, they have little participation in the program. It's almost like having two silent partners.

Hal: I think our state is kind of go do your own thing, and as long as we're not making waves or drawing any sort of negative attention that we're independently run.

Input Evaluation: Ideal Criteria

There were 11 ideal criteria for input evaluation in these motorcycle safety programs.

1. Amount of Funding. This criterion refers to amount of money and other resources available to conduct training programs. Administrators had little difficulty identifying how they would allocate additional resources if somehow more money became available. The level of funding dictated the number of students that could be enrolled into a rider education course, and administrators had as one of their primary objectives to have the program grow so more of the public could be served.

Carl: The defending for more funds would be real easy. We can simply show them where we are now and where we need to be. We could show them what we have and what we really need. Although, I think we do very well with what we've got, we could be much better.
Dan, in referring to how additional funds might be spent: *There’s tons; I would throw it at instructors, in their development... Get them some tools, some things they can use to improve our program; because, I’m convinced if the program works, it’s got to be working through the instructors.*

Edward: *Well, I think clearly one of the things we want to do is to lower the student tuition... I think tuition for some people is a deterrent. I mean, if you look at the people that are coming into the class, our average age is probably around forty, or maybe even forty-five... if we had more money, the first thing we’d be doing is looking at lowering the tuition.*

Greg: *The only thing the legislators look at every year, they come up to the money and they want justification, what they are getting for that money. So I have a one-page fact sheet for what they are getting for their dollars in order for them to approve the money every year.*

Hal: *I went to a conference, and I thought it was a real relevant statement; it was a representative form the National Highway Traffic Safety Administration that basically said there will be no additional funds given out to motorcyclists. And I admired his honesty, which is what we need more of. He said, in the grand scheme of things, if we have a certain number of children being killed due to a lack of car seats, based on “x” amount of motorcyclists being involved in crashes or fatalities, which one takes priority? Well, we’re not on the scale or the ladder of importance.*
2. Clear Communication As The “Measuring Stick” for Program. This criterion refers to the quality of communication within the program. There are several lines of communication, including administrator to her or his superior, administrator to program sponsors, administrator to instructors, and administrators to course participants. Overall administrators think it is important to be responsive to all stakeholders, and one of their primary functions is to gain feedback to make a program more responsive to the needs of all stakeholders.

Frank: Communications: it keeps my instructors aware of what's going on with the program and what we're doing and what we're looking to do in the future. We give them the numbers of what's happening, as far as training at the different sites. We throw in a little information on...every newsletter we throw in web sites related to motorcycling for their own personal use or they can pass on to the students in the class...That has been a strong point for my administration is communications. My instructors are very pleased with knowing what's going on. They don't feel like it's my program and I'm sitting up there dictating what we do and how we do it. They're getting constant feedback as to what's happening...it's greatly improved the unity among instructors.

Hal: My personal goal, of course, is to try to create an environment to where there is shared, open communication; no information ever hidden; there's no hidden agendas. And then trying to help each person understand where they fit into the process.
3. **Dedication of the People Involved.** This criterion refers to the instructors that teach the curriculum and sponsors that administer programs at the local level. Clearly, people were perceived as the primary resource for program operation. Since most administrators have statewide responsibilities, they must rely on the people involved at the local level. One of the interesting aspects of this criterion is that most administrators thought the more motorcycle enthusiasts (current riders) they had in key positions, whether above them as superiors or below them as site sponsors, the better the program operated. Non-motorcyclists were perceived as lacking the inherent motivation to assure that the program operated well.

Ben: *I don’t believe that we’ve had enough people in the state level and the government that’s been interested. Rather that’s someone that’s not really a motorcycle enthusiast, or they’ve strapped the budget down tight enough on funds.*

Carl: *And I also think that because we are so close-knit, we’ve got instructors that have been here a long time; they’re not afraid to call and make suggestions, which they do all the time which I may or may not be interested in. But at least they know that if they call and say ‘hey, there’s a problem here’ that we’ll try to address it; or ‘hey, I think we can do better doing this’; well, okay, let’s try it.*

Frank comments on a problem: *Instructor burnout is one. Instructors come in and…a lot of them come in for the money aspect. They try to work as many classes as they can to make money, and they forget why we’re really here and that’s for the love of the sport and trying to teach safe motorcycling. It’s still*
the goal but it's overshadowed by trying to make money. One too many classes and they start getting tired of it.

4. **Degree of Trust Within Program.** This criterion refers to the interaction between an administrator and her or his superior, and the interaction between instructors and course participants. Administrators expressed pride in their programs, and this feeling is closely associated with the degree of trust from their immediate superiors. Administrators were left alone because superiors trusted that the program and programmatic judgments were in good hands, or the level of communication was open to a degree that allowed program functions to operate unimpeded.

   Dan: *He (the supervisor) pretty much trusts me and he doesn't see to the day-to-day operation. He can't because he's over a whole section. I'm one of many functions that he oversees. He does, you know, get involved with complaints from time to time, and things that are brought to his attention.*

   Frank: *Simply because my immediate supervisor leaves me alone. It's my program. He let's me run it as long as everything seems to be running efficiently. He's there if I need him.*

5. **Dollars Spent On Quality Assurance.** This was also a currently-used criterion and was presented earlier with other currently-used criteria. The criteria that were both currently-used and ideal are presented in Table 3 on page 100.

6. **Having Adequate Budget To Meet Expectations.** This criterion refers to legislative and departmental allocation of funds for the motorcycle safety program. All
administrators could accomplish more with more money, and were quick to express how they would utilize additional funds. Demand for training programs is rising. The number of training courses that can be offered, and the amount of quality assurance measures that can be implemented are limited by resources available.

Frank: I think...and looking that the information may be negative, it might give us a little more fuel or ammunition to go before the legislative body that gives us our funding that the program needs to be more intensive to get results.

7. Having Adequate Number of Motorcycles. This criterion refers to the equipment needs for each of the training sites. Course participation is limited by the number of sponsors that offer a training program, that the number of sponsors is limited by the number of motorcycles available. Motorcycles are obtained either through dealer loan programs or by purchase.

Adam: ...it depends on the kind of money you're talking about. Soft money from the National Highway Traffic Safety Administration, we'd probably purchase some additional motorcycles and things for the program.

Carl: We have adequate personnel right now. Our biggest problem is not having motorcycles. If we had more motorcycles, we could do more training. We could run more programs at each site. We could do better on our expansion.

Hal: ...when it comes down to the actual concrete elements of equipment, we do have a need for particularly, more motorcycles, storage
buildings...Number-wise we are currently, based upon numbers right now, we're currently about sixty motorcycles short of what we need.

8. Maintenance Condition of Motorcycles. This was also a currently-used criterion and was explained with other currently-used criteria. This is presented in Table 3 on page 100.

9. People Skills of Administrators and Instructors. This criterion refers to the professional communication skills and human relationship skills of program personnel. Administrators are concerned with technical competence in delivering the curriculum, and with the quality of the instructor/participant interaction that occurs in the training program. These are the two primary issues referred to in quality assurance measures, particularly with administrators that conduct annual instructor updates.

Edward: I think one of the things we have is we need to, I think, tweak a little bit on instructor development. It's an area where we have instructors but we need to develop our staff a little bit better to help do those kinds of regular instructor development things; bring the level of instructor up. I think we need to do some sponsor development as far as working with the sponsors.

Hal: But the changes I would make have less to do with administrative structure as it does with administrative attitudes towards motorcycling in general, not the structure itself. The way we're set up the 'pecking order' that we have established or the hierarchy as it breaks down has worked effectively only because the people who are in those positions, you know, care about the program.
10. Quality of Instruction and Instructors. This was also a currently-used criterion and was presented earlier with other currently-used criteria. The criteria that were both currently-used and ideal are presented in Table 3 on page 100.

11. Training Site Coverage In State. The criterion refers to accessibility of citizens to training opportunities. The number of training sites and the distance from potential students is an important feature of the motorcycle safety program. Most programs are funded by state legislatures, and a primary goal of administrators is to assure that there is adequate state coverage. A hindrance is accessibility to a facility that has enough paved area for the on-cycle portion of the course.

Ben: I guess probably the primary objective is to expand our program to a statewide program with enough personnel and training sites and training equipment to do it with. That we get into state-owned property, which gives you more access and better facilities to work with. As a result, the main objective to this is to educate these people out here and see how many we get trained, how many get their legal license endorsement.

Carl: Our primary objective is to try to get everybody that rides a motorcycle in this state trained. That's our 'pie in the sky' objective. Our primary objective at this point, more realistically, is to try to identify and have training sites all over the state.

Frank: Biggest problem I have in _______ is finding suitable ranges. There's not a lot of good parking lot space. And that's one of the most difficult
functions that I have to do. With beautification going on around... we’re working on one site right now that has a local ordinance and a county ordinance; there has to be a tree within so many feet of every parking space. When you start planting trees in the parking lots, that makes it a little difficult for laying out ranges.

Process Evaluation: Ideal Criteria

There were 13 ideal criteria for process evaluation in these motorcycle safety programs.

1. Ability to Remedy/Remove Problems Within Program. This criterion refers to the power of a program administrator to control elements within a program. The issue is having a clearly defined process to control sponsoring agencies, and to remove troublesome instructors that do not teach the curriculum well or do not represent the program appropriately. Clearly, program administrators are involved at a level within a state system that is not significantly large, and their position of authority is within an agency that has other priorities.

Adam, commenting on lackadaisical sponsors: The biggest problem with most of them is either they didn’t know or they didn’t care. But we, by a lot of documentation that we have produced and sent out in the last year, we’re educating people and by personally seeing and interfacing with them, they’re getting to the point that they are going to change or they are going to leave. It’s easier to change.
Edward: *As far as any type of evaluation... if we see additional problems or maybe something that's unique at the site, then the field staff will try to set up a meeting with the sponsor to try to address that issue if there's something going on.*

2. **Capturing and Acting On Input From Instructors.** This was also a currently-used criterion and was presented earlier with other currently-used criteria. The criteria that were both currently-used and ideal are presented in Table 3 on page 100.

3. **Degree of Emphasis On Service Function.** This criterion refers to attitude of personnel within a program that focused on customer satisfaction. Service to the students is a primary focus for administrators. Administrators perceived their primary function to be of service to the instructors and program sponsors, and to assure that instructors were providing a service function for course participants. This criterion is closely associated with the communication criterion that also received a high priority within program operations. An additional function of this criterion was to assure that resources are utilized well, and that equipment was adequate in terms of supply and maintenance.

Dan: *The fact that we have expanded every year the numbers of people trained and, again, trying to focus on the service. They are getting a bang for their buck. That's not loading in the canon and shooting in the air. People are getting something back out of it.*
Frank in commenting on accountability said: *One is to make it available to everyone that wants to get into the course. Our goal is to try to make it available with a 50-mile radius.*

Hal: *Serving is the key word. We serve by first it's almost like a kingdom principle. It's all in reverse. If you want to be first, you've got to be last. If you want to reach a certain area in what you do, you have to become a servant first. So we serve by meeting the needs of those individuals, the needs of the program, and by using all those elements, those tools we have, we are able to steer, direct by serving the people.*

4. **Degree of Professional Development Among Instructors.** This was also a currently-used criterion and was presented with other currently-used criteria. The criteria that were both currently-used and ideal are presented in Table 3 on page 100.

5. **Extent of Constant Learning Within Program.** This criterion refers to the constant improvement among all the entities within a program. This is from top to bottom, from legislators to course graduates and the general public. Although the primary focus is improving the instructor interaction with course participants, this criterion also places emphasis on site sponsors. Administrators felt as program entities became more competent in each of their functions, they would then be able to address other areas of program operation that would improve the overall quality of the program and satisfy the intents of the program and assist in growth toward reaching more citizens.
Carl: I look at what other programs' ...kind of their evaluations when I go to SMSA conferences or something of that nature. I kind of listen to what they're doing and how they evaluate their programs, the numbers and percentages and all that. And I think we're doing real well. We have continued to increase the number of people we are training. We've increased the efficiency in which we've trained those people. And so I think that that's the way we do evaluations. I would love to have a formal evaluation procedure for instructors and all that.

The other thing that we do is at year's end, we sit down and we look at the numbers. We look at where they're coming from, and we see if we need to make any adjustments in our schedule.

Commenting on his own development, Dan said: I've taken a couple of classes and I've always picked up something from them. There's always things going on in that field (evaluation) that I'm not aware of. I feel that's really not motorcycle safety either. I beg for that kind of input too, when you talk about program evaluation. Because when you look at program evaluation, you're not just looking at motorcycle safety. You're looking at any kind of program and how those tools fit.

Frank: Every class that we run, I get a class report that comes in. I get student evaluations of the instructors and of the program. I look at every one of those when they come in. My administrative assistant puts all this information into a database and periodically throws out some charts or spreadsheets. She gives me something to look at. Like, for example, with this trip up here today, I'll
be on the road riding probably eight hours. And during that time, we (referring to his program assistant) are kicking around a lot of things as we go along. What’s happening with the program? What can we do to change things, to improve things? How can we hit more spots today? How can we work our schedules so that we can get more things done today?

Hal: To measure the effectiveness not only of our training but of the evaluation process, currently what’s in place is we have students doing an evaluation of the program and of the instructors. We have the administrative staff evaluating, doing observation visits on the instructors and the sites and the sponsors. We have our administration, my position and our staff members, being constantly evaluated by our project director.

6. Extent To Which Communication Is Open. This criterion refers to all entities within a program having a clear understanding of mission and goal achievement. Administrators expressed an appreciation for having supervisory feedback from the top and instructor feedback from below. The primary form of communication with superiors was formal meetings, and the primary form of communication with instructors is by telephone or email correspondence. Administrators also would liked to hear from program graduates, and several of them expressed pride in the fact that many course participants, when provided an opportunity to communicate with them, were very complimentary.

Hal: You know, communication means more than just the sharing of information. It means to actually understand the motivation of the person you are
talking to, and developing a tool that will allow us to understand where that person is coming from. And I feel in this program that we are not only...I was trying to keep away from saying this, but if we compare ourselves to other programs, I see us as being very successful in that aspect, because I can see even our instructor base...of other instructors I've talked to around the country, I see our instructors really in-tune and aware of moving toward focusing on the students...There's a sense of freedom there because if the instructors feel that way, then we're doing our job to allow them to ask those relevant questions.

7. Extent to Which Program Participation Is Fun. This criterion refers to the fact that motorcycling is intrinsically fun, and that any program affiliated with motorcycling should be a pleasant experience. This criterion refers to fun for course participants in that their training experience should be enjoyable, instructors in that there should be satisfaction and joy in teaching, local sponsors in that they should be proud because they provide a local service to the community, and administrators in that they oversee a strong and viable program.

Greg: *I think the strongest thing is that we've had good communication and the instructor quality has really increased over the last few years; that people are working and want to do it and they want to do it right; and, they want to have fun doing it at the same time.*

8. Extent To Which Quality Assurance Measures Are Implemented. This criterion refers to mechanisms within a program to improve the quality of the service. Not all quality assurance mechanisms require that money be spent. Administrators visit
courses as they can, and encourage instructors to help each other in teaching the course, and to assist program sponsors in maintaining the quality of local programming. For example, opportunities exist for instructors, particularly in areas where there is a cadre of instructors, to collaborate with each other and sponsors in such things as motorcycle maintenance, mentoring, sponsor assistance with registration, and dealer communication and outreach.

Adam: The only thing that hurts right now is that some of the quality...you know, the assurance measures are just not getting in place. For instance, for seven years I've been doing quality assurance reports, but I can show you all in my file cabinet where they went from me to the (former) coordinator and from the coordinator to the file cabinet and the loop was never closed.

Ben: And from time to time, if I'm not involved in the class and get a chance to drop in on the instructor teaching a class, I like to see the classroom presentation just a few minutes of it or the whole thing; same way on the range-to see if they're really looking at what they should be looking at and if they're grading what they should be grading...I try to feel like that by watching and by looking at these situations and assisting in any other way as far as a support person to the instructors that we can make the program better.

Frank in commenting on quality assurance provided an example of quality assurance without much cost: ...we look at pass/fail rates and drop rates within a class. We look at the statewide average...we can look that up on individual instructor basis, and we can look at it on an individual site basis. For example, I
had one site I was looking at recently and their drop or failure rate was about three times what the state average was. Of course that threw up a red flag, why is that happening? And I went in and looked at what they were doing different, and they were running the program a little differently than other people.

9. Extent To Which Quantity and Quality Are Raised Simultaneously. This criterion refers to the balance between expanding a program to generate more graduates, and improving the service at a current level. This is an all-encompassing criterion, and calls for astute judgment by program administrators. They realize, for example, that the overall goal of their statewide program is to serve the citizens, but at the same time assure that the quality of the service is adequate. From a practical viewpoint, this means to train more motorcyclists each year with minimal complaints from stakeholders.

Adam: Quality is more important than numbers, even though numbers drive a lot of outside agencies. If the quality isn’t there, the numbers are insignificant.

Dan: And again, we’ve kind of gotten away from looking at outcome as much as providing a service. That’s kind of what we try to look at and we need to continually improve on the service and find more efficient ways to deliver the service, and ways to improve the quality of it and that’s where we’ve gone to.

Edward: You can have the best quality program, you can have the best instructors, the best facilities and I think people are still going to take risks that may seem inconsequential at the time but can have major ramifications. And
that’s always a question I struggle with between quantity and quality. And the questions I struggled with from the very beginning is you want to have the courses available; yet, you also want to have quality programs at the same time. And the question tome has always been how you raise both at the same time.

10. **Level of Communication Within Program.** This criterion refers to the amount of communication that exists among program stakeholders. This differs slightly from the openness of communication in that as long as some communication occurs, it is an indicator of program activity. For instance, several program administrators think that having a newsletter is important so that instructors can feel that they are an integral part of the program. Other administrators prefer telephone or electronic communication; and, this may be a function of program size.

Frank: *I think that communications is, again, a big key. In talking with a lot of other administrators, or talking with instructors from other states, it seems that we have as good as communications as anyone, better that a lot. And I think that’s the key to any kind of organization is having communication, and that’s two-way communication. We’re not sitting up in an ivory tower anywhere. We’re pretty approachable by anybody about anything.*

11. **Quality of Instructor Updates.** This was also a currently-used criterion and was presented earlier with other currently-used criteria. The criteria that were both currently-used and ideal are presented in Table 3 on page 100.
12. **Quality of Interaction Between Instructors and Students.** This criterion refers to the connecting features of curriculum, instructor, and participant. This takes the form of accurately delivering the curriculum as designed, and doing so with professionalism and appropriate communicative style. Administrators stated that most of the complaints within a program arise because of the interaction between individual instructors and students.

Dan: *We get a tremendous amount of compliments. For the number of people we do put through, I hear...the complaints are usually petty...they are really program unrelated in some cases. In other cases, they're personality type issues...much more positive feedback concerning the program. They still are very enthusiastic about the experience, their instructors, what they learned.*

Greg: *Some of the students will e-mail me. I'm getting more and more e-mail for the students after class wanting to compliment the program, wanting to compliment the instructors. I used to get a lot of letters every year, may up to 20 letters written by students. Now I'm getting a lot of e-mail.*

14. **Way Problems Are Addressed.** This criterion refers to how a state administrator deals with major and minor problems that arise from program operation. These could be fiscal in nature, but mostly refer to dealing with the complaints of course participants and instructor complaints. This criterion is closely associated with “ability to remedy/remove problems within programs,” but is differentiated by the tone and attitude with which communication takes place. One of the difficulties in program operation is establishing quality assurance measures to promote good programming, and also to
maintain professionalism regarding how people are treated if their performance is less than adequate.

Dan: Well, like I told you, the feedback I get from students. Not only reading their evaluations but I field all the calls, the complaints. They all come to us. And all the good stuff comes to us too. I keep a file on complimentary letters on instructors and the program. Feedback from the instructors, I certainly get. If we've got problems with sponsors, I hear from them. Any problem they have in conducting class. We make a big point in making sure they know that we know. Otherwise, we can't help them out there as they do their job. So, we get a lot of feedback both from instructional staff and students. Get feedback from our public information campaigns as well.

Frank: I think my philosophy of management versus the previous director's management, one is basically...you try to hire the right people. And then you try to give them the right tools, the right training or funding or whatever it is they need to do their job, and let them do their job. I'm not dictatorial at all, and if you've got a problem and can't get your job done and need assistance, I'm there to assist you to do that.

Product Evaluation: Ideal Criteria

There were four ideal criteria for product evaluation in these motorcycle safety programs.
1. Course Student Evaluation Results. This was also a currently-used criterion and was presented earlier with other currently-used criteria. The criteria that were both currently-used and ideal are presented in Table 3 on page 100.

2. Increased Learning By Students. This criterion refers to the amount of knowledge and skill gained by course participants, and includes an analysis of knowledge test scores and skill test scores. An additional measure is customer satisfaction that is reflected by graduate surveys at the end of a course, or by informal feedback. One of the difficulties in assessing knowledge and skill gains is that participants are not assessed prior to program enrollment, and instructor evaluation of results is a primary source of determining increased learning.

Greg provides an experience of learning not necessarily captured by formal evaluations: 

"I have had people come to me and say 'look, this happened to me a couple of weeks ago and I remember that you taught me, and all I can say is that I'm glad I had the training because I wouldn't be here if I hadn't.' I've had instructors come to me and tell me, you know, some story like that. They are sitting at a restaurant eating with their wife or whatever, and someone will be staring at them, and they turn around to figure out why this guy is staring at them, and a few minutes later come over and introduce himself and say we are here today because of your training. We want to just thank you and let you know...it’s really hard to evaluate that and give a true reality value to how much we are worth."
Hal commented on another way to assess increased learning: I'd like to see us go more towards pre- and post-testing. You know, which we do the post but not so much the pre-testing. I'd like to see the participants do a self-evaluation, the students, do their own evaluation. I'd like to see the students evaluated by the instructors. We do that per se according to the skills test but I find that to be very shallow. We're gauging skills of riding a motorcycle based upon the knowledge of the instructor based upon what he or she was supposed to have helped the student learn. So I'd like to see an actual evaluation done by the instructor including classroom and the person's personality.

3. Increased Skill By Students. This criterion refers to the skill of course graduates as measured by an end-of-course skill test. Administrators pay close attention to the skill test scores and pass/fail rates as an indicator of course and instructor quality. There is a perception that increased skill is a primary outcome of the course, and this appropriate since the focus on motorcycle safety training is on skill development.

Ben: I'm encouraging those individuals (course graduates) to stay in contact, write letters or email or whatever, and let us know what their experiences are as they continue to improve their skills; and I get a lot of that, a lot of literature and feedback.

Edward: I mean, it's always hard to quantify a negative. How can you say that someone didn't get in an accident, someone didn't get injured, someone didn't do something? It's hard to do that. I think you can do some by research and you can do some by anecdotal. But I think if you see the course and you see
what the instructors are teaching, see the curriculum being administered, I think we’re clearly saving lives.

Frank: Our ultimate goal is to put people out there that are as safe a motorcyclist as possible. These people are taking this course because they want to learn about a motorcycle. And I’d be foolish if I believed that just because they didn’t complete the course, that they’re not going to go out there and ride a motorcycle anyway. So, in meeting our goal, yes, we need to give every possible chance to get that student through and get them to learn the skills.

4. Training Itself Is Safe. This criterion refers to the number of accidents and incidents that occur during the training program. Administrators acknowledged that an overall purpose of their program was to reduce the number of motorcycle deaths and injuries. But a feature they have some control over is the safety of the training itself. They expressed that a measure of program quality is the safety of the training itself, and not what graduate experience was after successful course completion.

Adam: I also keep the number of training accidents at the sites. I keep...they’re required if it’s above a band-aid, they’re required to submit the report to me...Because that is also an indicator, and this program is not supposed to have accidents. Accidents occur we know that, but it’s not supposed to generate accidents during the training of students, normally. It’ll happen every now and then but it’s not something that should be a continual occurrence.
Carl: I think it should be held accountable in that we are a statewide program. I think it should be held accountable that when we bring people in, we train them in a very safe manner. And it's done in a professional manner, in a non-threatening manner, and it's done by people who are trained.

Part Three. Comparing Currently-used and Ideal Criteria

Seven criteria, shown in Table 3, were both currently-used and criteria identified as ideal.

Relative to the CIPP model, there were no criteria related to context evaluation, three criteria related to input evaluation, three criteria related to process evaluation, and one criterion related to product evaluation.

Table 3

Criteria Both Currently Used and Ideal For Program Evaluation

<table>
<thead>
<tr>
<th>Context</th>
<th>Input</th>
<th>Process</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>Dollars Spent on Quality Assurance</td>
<td>Capturing and Acting on Input from Instructors</td>
<td>Course Student Evaluation Results</td>
</tr>
<tr>
<td></td>
<td>Maintenance</td>
<td>Degree of Professional Development Among Instructors</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Condition of the Motorcycles</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quality of Instruction and Instructors</td>
<td>Quality of Instructor Updates</td>
<td></td>
</tr>
</tbody>
</table>
Ten criteria, shown in Table 4, were currently used but not identified as ideal. Relative to the CIPP model, there were no context evaluation criteria, no input evaluation criteria, seven process evaluation criteria, and three product evaluation criteria.

Twenty-three criteria, shown in Table 5, were identified as ideal but not currently used. Relative to the CIPP model, two were context evaluation criteria, eight were input evaluation, ten were process evaluation criteria, and three were product evaluation criteria.

Chapter Four presented the findings of this study. It provided the 17 currently-used criteria and 30 criteria identified as ideal. The seven criteria that were both currently-used and ideal were also presented, along with the criteria currently used but not ideal and the ideal criteria not currently used.
<table>
<thead>
<tr>
<th>Context</th>
<th>Input</th>
<th>Process</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>None</td>
<td>Capturing And Documenting Unsolicited Responses From Students</td>
<td>Accomplishment Of Pre-determined Goals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Degree Of Assessing Individual Sites, Not State As A Whole</td>
<td>Numbers Trained</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Degree of Formal Documentation Of Quality</td>
<td>Pass/Fail Rates In Courses</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Frequency Of Instructor Updates</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Having Student Waivers Upon Course Completion</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Requiring That Student Evaluations Be Completed</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Uniformity of Course Reporting</td>
<td></td>
</tr>
</tbody>
</table>
### Table 5

Ideal Criteria Not Currently Used For Program Evaluation

<table>
<thead>
<tr>
<th>Context</th>
<th>Input</th>
<th>Process</th>
<th>Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority Of Agency In Which Program Is Conducted</td>
<td>Amount of Funding</td>
<td>Ability To Remedy/Remove Problems Within Program</td>
<td>Increased Learning By Students</td>
</tr>
<tr>
<td>Support From Superiors</td>
<td>Clear Communication As A Measuring Stick</td>
<td>Degree Of Emphasis On Service Function</td>
<td>Increased Skill By Students</td>
</tr>
<tr>
<td></td>
<td>Dedication Of People Involved</td>
<td>Extent Of Constant Learning Within Program</td>
<td>Training Itself Is Safe</td>
</tr>
<tr>
<td></td>
<td>Degree Of Trust Within Program</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Having Adequate Budget To Meet Expectations</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Having Adequate Number Of Motorcycles</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>People Skills Of Administrators And Instructors</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Training Coverage In State</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5 (Continued)

<table>
<thead>
<tr>
<th></th>
<th>Level Of Communication Within Program</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quality Of Interaction Between Instructors And Students</td>
</tr>
<tr>
<td></td>
<td>Way Problems Are Addressed</td>
</tr>
</tbody>
</table>
CHAPTER 5

SUMMARY AND CONCLUSIONS

This chapter contains a summary, discussion, conclusions, and implications of this research. Currently used and ideal program evaluation criteria are discussed within the framework of the Context-Input-Process-Product (CIPP) model of program evaluation (Stufflebeam, 1971). Implications for research and practice relating to motorcycle safety rider education and training programs are drawn from the findings of this study.

Summary

The purpose of this study was to identify program evaluation criteria for motorcycle safety rider education and training programs from the perspective of state program administrators. The two research questions were “What are the currently-used criteria for program evaluation in motorcycle safety rider education and training programs?” and “What are ideal criteria for program evaluation in motorcycle safety rider education and training programs?” Eight administrators were interviewed using questions developed from the framework of the CIPP model of program evaluation. A total of 205 evaluation criteria were initially identified from the analysis of the interview results data, and administrators named 17 of these criteria as currently used for program evaluation. Through the use of a double-round Delphi technique, these administrators rated 30 criteria as being ideal for program evaluation.
Discussion

The identification of 205 criteria for consideration in program evaluation indicated the depth and breadth of motorcycle safety rider education and training programs. The complexity and scope of operations of these programs require state administrators to identify and monitor a wide range of criteria, in terms of which the quality of their programs is assessed. These criteria relate to human, material and financial resources at several levels of the system involved in motorcycle safety rider education and training programs. However, only 17 of these criteria were identified by one-half or more of the administrators interviewed as being currently used in their programs. It is likely that many of the 188 criteria were redundant and others may not have been recognizable by all administrators when these criteria were presented in the first round of the Delphi probe. Assuming that the majority of criteria were distinctive and identifiable by all administrators polled, the net results showed a wide variety of viewpoints concerning evaluation criteria, and the results indicate that no consensus exists in terms of the majority of evaluation criteria initially identified in interview results. Similarly, only 30 of the 205 criteria were identified as having “more value” by one-half or more of the administrators during the second round of a Delphi probe. Following is a discussion of each set of criteria and comparisons between the two sets.

Currently-Used Criteria

Of the 17 currently-used criteria, the largest single group (n=10) identified by program administrators was in the category of process evaluation. Process
evaluation provides information about the strengths and weaknesses of a program strategy during implementation so that the strategy or its implementation might be strengthened. An examination of the ten criteria cited as currently used in motorcycle safety rider education and training programs shows a concern for gathering information about instructor feedback, such as instructor updates of their teaching activities, instructor assessment of local programs, and uniform reporting procedures by instructors. Process criteria also relate to student activities, such as completed student evaluations, the use of unsolicited responses from students, and such specific requirements as skill test waivers given to students upon course completion. All of these items were local in nature and concerned assessment at individual course sites. None reflected a concern with statewide program processes. These criteria indicate that state administrators were focused on gathering information that identified strengths and weaknesses in the delivery of on-site instruction.

Three input evaluation criteria were in current use. Input evaluation provides information about the strengths and weaknesses of alternative strategies for achieving program objectives. This includes information about such program features as capabilities of the responsible agency, program design, and the variety of resources used to operate programs. The criteria identified by administrators as currently used relate to both human and material resources. More specifically, the criteria indicated a concern for funds spent to ensure quality of programs and facilities. As in the case of process criteria, these input criteria reflected an orientation to local programs and local resources.
The four product evaluation criteria also indicate a local program orientation. Product evaluation refers to determining whether objectives are achieved and whether the procedures 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 ... standards, and making rational interpretations of the outcomes using the recorded context, input, and process information" (Stufflebeam, 1971, p. 232). The four criteria in current use meet certain aspects of this definition of product evaluation, but they are lacking in other aspects. The criteria focus on specific, local and largely quantitative indicators of program outcomes, including numbers of students trained, pass/fail rates, and student evaluation results. As might be expected, administrators emphasized the accomplishment of predetermined goals. For the most part, administrators identified these goals in terms of fixed requirements of the national, standardized curriculum of the Motorcycle Safety Foundation, and most of the administrative procedures were implemented by replicating other state programs and by standards set by the resources of the National Association of State Motorcycle Safety Administrators. There were state-specific idiosyncrasies, but the primary feature of all these programs was the operation of the basic 20-hour curriculum.

The product evaluation criteria were also closely related to the input and process criteria described earlier, especially in terms of their orientation to local program activities and their attention to student and instructor activities. However, the administrators did not identify context evaluation criteria in current use.
Administrators indicated that their programs should only be held accountable to the degree the program met the letter of the law. The law requires only that training be available to the general public. A few administrators stated that it was important that they met their supervisor's expectations, while others stated the programs' ultimate goal was to reduce crashes and injuries. This left unclear exactly what the administrators were evaluating, in terms of context with which their programs operate.

Overall, the currently-used criteria reflect a pragmatic, local orientation by administrators as they gather information about their motorcycle safety rider education and training programs. Lacking on this list of 17 criteria was the specific attention to statewide program concerns. It is possible that administrators considered the curriculum and funding decisions to be out of their immediate influence, and thus were not likely to attend to them in routine program evaluations. However, they were not unconcerned about context evaluation, as indicated in the findings that relate to ideal evaluation criteria.

Ideal Criteria

Of the 30 ideal criteria identified as ideal, two were context evaluation criteria. Although their foremost responsibility was to follow the letter of statutory legislation and administrative regulation in their respective states, administrators identified "priority of agency in which program is conducted," and "support from superiors," as ideal criteria. These two context evaluation criteria indicate that more attention ought to be given to the present framework within which the program operates. For instance, the criterion "Priority Of Agency In Which Program Is
Conducted” showed that these administrators considered how a program was situated in a state department agency might influence funding decisions and program longevity. Some administrators interviewed indicated that they would like to have their personal job performance, as well as program outcomes, evaluated in terms of the program’s priority within the state system. In other words, their programs should be evaluated with consideration given to the resources allocated. In the interviews, some administrators did not think that their program received due recognition from their superiors.

The largest single group of criteria (n=13) identified by program administrators was in the category of process evaluation. The administrators again focused on local programs and on-site instructional and student activities. These criteria were under the direct control of administrators, and they reflect the perceived importance of how programs operate on a daily basis. For example, the criterion “Extent Of Constant Learning Within Program” was an indication that constant improvement was important to these administrators, particularly in terms of their day to day decision making. These process evaluation criteria also highlight the service orientation of administrators, to include the quality of services extended to students. Administrators thought the quality of a program should be connected to the service that was provided.

The next largest group of criteria (n=11) was input criteria. These focused on personnel, budget, facilities and equipment, as did the currently used criteria. The criteria demonstrated administrator interest in the degree to which personnel were trained and the quality of equipment that is available to students. However, the ideal
criteria also included items related to communication and the need to provide statewide coverage in terms of training programs offered by the state agency. The input criteria were indicative of the resources that administrators were concerned about the most. These criteria showed that administrators were interested in resources to the extent that the needs and interests of potential participants were accommodated. Besides having sufficient resources, these criteria indicated that trust and communication were also important. For instance, the criterion “People Skills Of Administrators And Instructors” indicated that interactions of people within the system were important.

Input evaluation criteria indicated that administrators would like to have more emphasis on budget allocation, communication, and the dedication of key program personnel such as program sponsors and instructors. Clearly, these administrators thought that resources available affected program outcomes, and to evaluate a program without consideration for expenditures was inappropriate.

The four ideal product evaluation criteria were directly related to the local program outcomes. These criteria mainly addressed student outcomes and the safety of training programs. Administrators considered increased learning and increased skill as important, and they would like to have methods to measure these indicators of program effectiveness. The administrators felt that motorcycle safety rider education and training programs should be a portal leading to a safer motorcycling environment.

Perhaps the most compelling ideal criteria were in the product evaluation category. These were “Increased Learning By Students,” “Increased Skill By Students,” and “Training Itself Is Safe.” The criterion that related to the training itself
being safe could become an important intermediate measure over which administrators have a degree of influence and control.

The identification of 30 ideal criteria for program evaluation demonstrated that these administrators had a clear systemic view of program operations, and given the opportunity to assist in the development of a broader evaluative framework, administrators would be able to approach quality programming from a variety of stakeholder perspectives. Although administrators indicated that they would like to have ongoing, objective program evaluations conducted that would help them improve their management functions, none of them felt qualified to formally evaluate the overall program in a systematic way.

It is important to note that only two ideal criteria were rated unanimously as being more valuable than the others. They were “Quality Of Instruction And Instructors” and “Quality Of Instructor Updates.” Clearly, these administrators considered the essence of their program to be the quality and performance of the instructors, and placed a high priority on overall professionalism and teaching effectiveness. Instructors were the focus of program improvement efforts, and aside from the allocation of additional resources in order to expand training, their professional development was paramount.

Additionally, communication was a significant indicator in the ideal criteria. It was named three times in different contexts: clarity of communication, level of communication, and openness of communication. If these administrators perceived one primary duty over which they have control, it would be to communicate clearly at all levels within their organizational structure, and to maintain an open, honest
environment. They strongly believed in the value of their programs and were
unafraid to publicly acknowledge their efforts.

Comparing Currently-used and Ideal Criteria

Additional meaning of the findings can be gleaned by comparing actual
criteria to ideal criteria. The criteria can be viewed from three perspectives. One is to
examine the seven criteria that were both currently used and identified as ideal; a
second perspective is to examine the ten criteria that were currently used but not
identified as ideal; and a third perspective is to examine the twenty-three ideal criteria
that were not currently used.

Currently-used and Ideal

Of the seven criteria that were both currently used and ideal, three were input
evaluation criteria, three were process evaluation criteria, and one was a product
evaluation criterion. There were no context evaluation criteria that appeared on both
lists. Overall, these seven criteria relate mainly to instructors and the quality of their
work. The criteria also address funds spent on quality assurance and maintenance of
equipment. These administrators think that the heart and soul of their program is the
performance of their instructors, and have a strong desire to ensure that facilities are
in place to support the quality of work done by instructors.

Currently Used But Not Ideal

Of the ten criteria that were currently used but not identified by administrators
as ideal, there were no input evaluation criteria, seven were process evaluation criteria
and three were product evaluation criteria. The process evaluation criteria relate to
student and instructor data and procedures. (As discussed earlier in this chapter,
administrators cited no currently-used context evaluation criteria.) This difference in currently-used and ideal input criteria may be attributed to the fact that administrators have little influence regarding resource allocation. While administrators desire additional monies for their program, they typically communicate this informally with immediate supervisors. Their primary focus is to efficiently and effectively utilize the resources that are provided from year to year.

Three currently-used process evaluation criteria stand out in terms of their potential as ideal criteria. The criterion “Capturing And Documenting Unsolicited Responses From Students” would appear to be a powerful criterion of program worth. This currently used criterion received the most emphasis by administrators during the interviews. Unfortunately, there were no formal procedures to document and report student comments after graduation. But such information has the potential to provide important evaluative information. Of note is that “Frequency of Instructor Updates” was not named as an ideal criterion. Since the quality of instruction was a major concern of administrators, it would appear that more frequent professional development opportunities would be embraced. The third criterion, “Having Student Waivers Upon Course Completion,” related to how these motorcycle safety programs interfaced with a state’s licensing entities. One of the more important values of these education and training programs was to make the licensing process more efficient for participants. Having a license waiver upon successful course completion would be an example of a value-added feature, and it would provide an opportunity to demonstrate a connection with another state agency.
It is also surprising that three currently-used product evaluation criteria, “Accomplishment Of Pre-Determined Goals,” “Numbers Trained,” and “Pass/Fail Rates in Courses” were not identified as ideal criteria. All three could be primary indicators of program efficiency and effectiveness. It could be they were not ideal criteria because program administrators had no direct control over the program elements reflected in these criteria.

Ideal Criteria Not Currently Used

Twenty-three criteria were identified as ideal, but were not currently used in program evaluation. These criteria included two criteria for context evaluation, eight criteria for input evaluation, 10 criteria for process evaluation, and three criteria for product evaluation. One feature of this list is the inclusion of two context evaluation criteria. Another is a balance in the number of process and input evaluation criteria, and the third is the qualitative aspects of criteria included on the list.

The addition of context evaluation criteria has been discussed, but the criteria stand in sharp contrast to the lack of such criteria in other comparisons. Although administrators appear to be uncertain of their current ability to evaluate agency priorities and the support afforded them by their supervisors, they seem willing to include such evaluations under ideal circumstances.

More input criteria are listed that appear among the 17 criteria that are currently used for program evaluation (described earlier in this chapter). The additional input criteria deal mainly with such “soft” features of program operation as communication and “people skills.” There is also an additional concern for ensuring statewide coverage in programming.
The process and product criteria reflect an overall concern with quality and a greater depth of concern for some of the same factors listed in the currently used criteria. Whereas the emphasis in currently used criteria is on metrics (e.g., number of students trained and frequency of instructor updates), the additional ideal criteria speak to such factors as "constant learning", "extent to which quality and quantity are raised simultaneously" and the "quality of interaction between instructors and students." One implication of this comparison is that, given ideal conditions, administrators would prefer to "dig deeper" into the dynamics of their programs and examine them for qualitative indicators of operations and outcomes. Their focus remains on students and instructors, but ideally, they would opt for an evaluation system that provides more than what routine indicators of program effectiveness afford.

Overall, the list of criteria not currently used for program evaluation suggests that there is considerable opportunity for improving program evaluation; that is, the inclusion of these criteria could substantively enhance program evaluation efforts if administrators are able to extend their currently used criteria to include criteria from all four categories of the CIPP model and extend their focus on process and input criteria to include additional qualitative indicators of program effectiveness.

Conclusions and Implications

State motorcycle safety rider education and training programs have a multitude of criteria from which an approach to program evaluation can be developed. Identification of over 200 evaluation criteria illustrates the variety and
scope of criteria available to administrators whose decision making can be served by the use of a systematic approach to program evaluation.

There are qualitative and quantitative differences in currently-used criteria and the ideal criteria for program evaluation in motorcycle safety rider education and training programs. The results of the study provide a basis for further analysis of the value both sets of criteria have for decision making at the state level of program evaluation. For example, some currently-used criteria may be targeted if deemed important for program evaluation. There are some ideal criteria that may not be easily adopted by state administrators due to budgetary constraints or other program limitations. However, an assessment of criteria in terms of their worth as aids to decision making is a recommended first step toward development of a relevant and practical list of criteria for program evaluation.

The results of this research suggest that evaluative efforts should be attentive to three interactions within these programs. These are: 1) The interaction between program administrators and their superiors, 2) The interaction between the administrators and the instructors, and 3) The interaction between the instructors and the students. These interactions could form the centerpiece of future program evaluation initiatives.

The first interaction concerns the interfacing between program administrators and their superiors. Program administrators had mixed reactions regarding their relationship with their supervisor, but nonetheless were concerned with how superiors assessed their programs. That administrators named "Priority Of Agency In Which
Program Is Conducted” and “Support From Superiors” as the only two ideal context evaluation criteria shows the significance of this interaction.

The second interaction concerns the relationship between administrators and the instructional staff. Although they did not have frequent personal communication with instructors, administrators put substantial efforts into assuring the quality of instructors and their teaching activities. Since the administrators considered the instructors to be the focal point of program effectiveness from a customer satisfaction viewpoint, the caliber of instructor and the degree of their professional development were priorities. Most of the process evaluation criteria were related to instructor issues.

The third interaction concerns the interaction between instructors and students. The quality of the education and training program lies in this relationship, and the student centered ideal product evaluation criteria indicate the importance placed on the student experience at the local level.

Product evaluation has the most compelling criteria that could be addressed in program evaluation activities. Increased learning and skill development are direct results of motorcycle safety rider education and training programs, yet they are not currently used as evaluation criteria. The third criterion listed relates to the training itself being safe and could become an important intermediate measure over which administrators have a degree of influence.

It was evident that program evaluation was an ongoing concern for administrators of these motorcycle safety rider education and training programs.
They consider evaluation an ongoing process. They are more interested in evaluating the effects of each decision they make than they are with scientific evidence of the safety value of education and training. Their perspective regarding program evaluation is to expand the program to meet public demand and to ensure that a quality experience is provided to program participants.

This research raises additional questions for program evaluation in motorcycle safety rider education and training programs: 1) Should these programs be evaluated for the value added in knowledge and skill gains or on their effect in reducing national crash and fatality rates? 2) Should program evaluation efforts focus on program improvement or on crash and fatality rates? 3) Should studies of crash and fatality rates be the determining mechanism to determine the worth and value of these motorcycle safety and training programs?

Regarding the first question, there are two primary themes in motorcycle safety literature regarding program evaluation. One is that quantitative studies have yet to find the methodology that proves these training programs work. Second, there has been a call for some way to prove the efficacy of such programs, usually motivated by the fear that programs will be eliminated if they don’t prove their worth in reducing deaths and injuries. Clearly the safety training community is perplexed by the inability to demonstrate value, and proving gains in knowledge and skill would put these programs in a defensible position.

The National Agenda for Motorcycle Safety (National Highway Traffic Safety Administration and the Motorcycle Safety Foundation, 2000), points out that “The effectiveness of the concepts discussed in this document require a foundation of
viable and current research in most areas pertaining to motorcycle safety. While there is a substantial body of work relating to motorcycle safety in the United States and abroad, few of these studies, research projects, or statistical reports were done in coordination with one another” (p. 7). This document calls motorcycle rider education the centerpiece of a comprehensive motorcycle safety program, and states the challenges are “to get motorcyclists to take training and to keep quality rider training affordable and accessible to all interested parties” (p. 17). Three recommendations that were named regarding rider education and training were: 1) Expand motorcycle safety programs to accommodate all who need or seek training, 2) Conduct uniform follow-up research into the effectiveness and impact of rider education and training, and 3) Establish benchmarks for rider education and training effectiveness and program operation excellence.

Regarding the second question, the expectations of evaluation are unclear: on one hand evaluation is to prove effectiveness, and on the other hand it is to improve the program. The fact over 200 criteria were initially identified from this research is evidence of the complexity of these programs. It also points out that although program evaluation is not a formal process, it is a primary concern of program administrators. Administrators interviewed in this study freely provided their thoughts, and they desired a credible mechanism from which to demonstrate program worth. They seemed to be searching for a way to prove that they were accountable for program quality and improvement. They were interested in evaluation methods that would end speculation regarding the question of program effectiveness. These
administrators desired clean, unambiguous answers to questions surrounding program effectiveness and efficiency.

Regarding the third question, a focus on crash and fatality rates diminishes the perceived value of these programs' efforts. Training and education programs should not bear the full responsibility for problems that are not within its realm of effecting behavioral change that affects rider choices beyond a course. History shows a pattern regarding safety training programs and their evaluation. This pattern begins with a spike in crash statistics and fatality rates, followed by a public outcry that something needs to be done. Monies are allocated to implement safety countermeasures that usually revolve around training and education activities. The intent is to attenuate the increase in crash and fatality rates. Training and education, it is hoped, will cause a swift and pervasive social change in unsafe behavior. To prove the benefits of a safety training program, the evaluation community calls for an evaluation of the crash and fatality reduction outcomes. Programs are diffuse, funding is unstable, and there is uncoordinated multiplicity of programs. So, research studies commence in an attempt to connect training and education with a reduction in crash and fatality rates. However, their results show no differences in crash or fatality rates. A call for program elimination follows until the next spike or crisis is uncovered or publicized.

The motorcycle safety education and training community in particular and the traffic safety education community in general have serious concerns about the evaluation of safety training programs. Using broad social indicators to determine the worth of a program could be an inappropriate application of evaluation practices. The value of a program should be determined by using intermediate indicators; in other
words, motor vehicle safety programs should be evaluated not by crash and fatality rates, but rather by increases in the knowledge and skill of program graduates. This was alluded to recently by Driver/Education (2000): “Driver/Education has always maintained that reduced crash risk is an elusive mission for the training/education community to set out to accomplish, and that is counterproductive to set reduced crash rates as the only reason to train and educate drivers. As a mission, this channels research and creative energy towards a goal that is extremely difficult or impossible to achieve. It squeezes out goals and objectives that are more real to drivers’ everyday lives and that might, indirectly, be more effective in reducing crashes (p. 9). This concern places the focus back onto the programs themselves, as they can be evaluated in terms of their internal and contextual features. It places responsibility for evaluating programs in the hands of administrators with overall responsibility for these programs.

A partial solution could be to utilize intermediate measures, particularly as they relate to product evaluation. Suggested criteria resulting from this research were: 1) Participant satisfaction, 2) Gains in knowledge, 3) Gains in skill, 4) Training itself is safe, 5) Graduate input after having some experience which captures testimonials, and 6) Instructor satisfaction.

A realistic, acceptable, and worthy model of program evaluation for motorcycle safety rider education and training programs should be developed. The ideal criteria for program evaluation identified from this research can serve as a beginning step toward the development of a viable model. These criteria were from the perspective of program administrators, and although evaluation was inherent in
their programmatic decisions, no formal mechanisms exist to evaluate efficiency and effectiveness. However, these criteria were identified by eight administrators in one region of the country. This study ought to be replicated in one or more other regions of the country in order to ascertain whether administrators in those regions would hold similar perceptions of ideal criteria.

Further research needs to be conducted to explore the basis of observed differences in ideal criteria and those in current use in state programs. In the present study, the researcher speculated on some of the reasons for differences in ideal criteria and those in use, and in many aspects the criteria seemed to differ only in scope and depth. Further investigation on this phenomenon could take the form of interviews with state administrators, in which state administrators are asked to discuss the barriers to implementing such ideal criteria in their systems.

Further research along these lines holds promise, in terms of its potential contribution to the implementation of a comprehensive and thorough evaluation system for motorcycle safety rider education and training programs. Such an evaluation system can result in improvements in safety programs and their efficacy can be demonstrated to the public and to policymakers. To the extent that improved programs result in greater safety of the program participants, such efforts will prove their worth.
References


Appendixes
Appendix A

Evaluation Criteria: Round One with Frequency of Responses
<table>
<thead>
<tr>
<th>Criterion</th>
<th>More Value</th>
<th>Of Value</th>
<th>Less Value</th>
<th>Cannot determine</th>
<th>Presently Used for Program Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to handle stand-by students</td>
<td>11</td>
<td>111</td>
<td>11</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Ability to handle walk-in students</td>
<td>11</td>
<td>11</td>
<td>111</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Ability to remedy/remove problems within program</td>
<td>1111</td>
<td>111</td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Access to resources</td>
<td>111</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accident and fatality rates</td>
<td>1</td>
<td>111</td>
<td>111</td>
<td></td>
<td>111</td>
</tr>
<tr>
<td>Accomplishment of pre-determined goals</td>
<td>1111</td>
<td>111</td>
<td>1</td>
<td></td>
<td>1111</td>
</tr>
<tr>
<td>Addressing galvanizing questions (like helmet law)</td>
<td>1</td>
<td>111</td>
<td>111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adequacy of finding emerging problems</td>
<td>111</td>
<td>111</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amount of advertising</td>
<td>11</td>
<td>11</td>
<td>111</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Amount of funding</td>
<td>11111</td>
<td>1</td>
<td></td>
<td></td>
<td>111</td>
</tr>
<tr>
<td>Amount of Instructor &quot;bellyaching&quot;</td>
<td>1</td>
<td>111</td>
<td>11</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Amount of time administrator is free of procedural duties</td>
<td>11</td>
<td>111</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amount of volunteer efforts within program</td>
<td>1</td>
<td>1111</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assuring some riders discover motorcycling is not for them</td>
<td>11</td>
<td>111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Availability and use of use of 402 funds</td>
<td>1111</td>
<td>111</td>
<td>1</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Balancing quantity of students and quality of program</td>
<td>1111</td>
<td>111</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Belief within program that motorcyclists must take care of their own</td>
<td>1111</td>
<td>111</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Capturing and acting on input from Instructors</td>
<td>11111</td>
<td>11</td>
<td></td>
<td></td>
<td>1111</td>
</tr>
<tr>
<td>Capturing and documenting unsolicited responses from participants</td>
<td>1111</td>
<td>1111</td>
<td></td>
<td></td>
<td>11111</td>
</tr>
<tr>
<td>Centralized administration</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Centralized registration system</td>
<td>11</td>
<td>1111</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>---------------------------------</td>
<td>----</td>
<td>------</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Clarity with which quality is defined</td>
<td>11</td>
<td>1111</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clear communication of the &quot;measuring stick&quot; for program</td>
<td>1111</td>
<td>11</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compliments to complaints ratio</td>
<td>111</td>
<td>111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conducting IPs only with CIs in state</td>
<td>1</td>
<td>1111</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Coordination of individual training sites</td>
<td>1</td>
<td>111</td>
<td>1</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Coordination of sites and course locations</td>
<td>11</td>
<td>111</td>
<td>11</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>Cost per rider</td>
<td>11111</td>
<td>1111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course student evaluation results</td>
<td>1111</td>
<td>111</td>
<td>11111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Courses produce learning in a Controlled environment</td>
<td>111</td>
<td>11</td>
<td>1</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Decentralized administration</td>
<td>11</td>
<td>11111</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dedication of the people involved</td>
<td>1111</td>
<td>111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree of assessing individual sites, not state as a whole</td>
<td>1111</td>
<td>1111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree of dealer involvement in program</td>
<td>1111</td>
<td>111</td>
<td>1</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Degree of distribution of information</td>
<td>1111</td>
<td>11</td>
<td>11</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Degree of emphasis on service function</td>
<td>1111</td>
<td>11</td>
<td>1</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Degree of financial flexibility afforded administrators</td>
<td>1111</td>
<td>111</td>
<td>1</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Degree of follow up on riders that complete courses</td>
<td>111</td>
<td>11</td>
<td>11</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Degree of formal documentation of quality</td>
<td>1111</td>
<td>1111</td>
<td>1</td>
<td>1111</td>
<td></td>
</tr>
<tr>
<td>Degree of informal peer pressure among Instructors</td>
<td>1111</td>
<td>111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree of Instructor control</td>
<td>11111</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree of meeting demand for training</td>
<td>1111</td>
<td>1111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree of motorcycle enthusiast involvement</td>
<td>11111</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree of professional development among Instructors</td>
<td>1111</td>
<td>111</td>
<td>1111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree of public support</td>
<td>11111</td>
<td>1</td>
<td>1</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Degree of reaching out beyond training courses</td>
<td>1111</td>
<td>111</td>
<td>1</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Degree of scrutiny of program (less being better)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Degree of stakeholder feedback</td>
<td>1111</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Degree of trust within program</td>
<td>11111</td>
<td>11</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Degree of use of technology</td>
<td>1</td>
<td>1111</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Degree to which administering agency understands program</td>
<td>1111</td>
<td>11</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Degree to which best practices from other states are adopted</td>
<td>11</td>
<td>1111</td>
<td>11</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Degree to which change is accepted</td>
<td>1111</td>
<td>1111</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree to which course graduates crash</td>
<td>1</td>
<td>1111</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Degree to which licensing and education dovetail (complement one another)</td>
<td>1</td>
<td>1111</td>
<td>11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Degree to which lives are saved</td>
<td>1111</td>
<td>11</td>
<td>1</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Degree to which local sponsors have control</td>
<td>1</td>
<td>11</td>
<td>111</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Degree to which program environment is apolitical</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Degree to which there is “shared vision” within program</td>
<td>11</td>
<td>1111</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Demand exceeds supply</td>
<td>1111</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Direct and/or indirect involvement in lobbying efforts</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Diversity of sponsoring agencies</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Documentation of year-to-year improvements exist</td>
<td>1111</td>
<td>111</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dollars spent on quality assurance</td>
<td>11111</td>
<td>11</td>
<td>1</td>
<td></td>
<td>111</td>
</tr>
<tr>
<td>Ease of attracting ICs</td>
<td>11</td>
<td>1111</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equipment has maximum utilization</td>
<td>1111</td>
<td>1111</td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Evaluation occurs by advisory committee</td>
<td>1</td>
<td>1111</td>
<td>11</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Evaluation of only those items that can be controlled</td>
<td>11</td>
<td>1111</td>
<td>111</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Extent of constant learning within program</td>
<td>11111</td>
<td>1</td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Extent of Instructor recognition</td>
<td>11</td>
<td>1111</td>
<td></td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Extent of post-course follow up</td>
<td>11</td>
<td>11</td>
<td>1</td>
<td>11</td>
<td>111</td>
</tr>
<tr>
<td>Extent to which attitude of service is maintained</td>
<td>1111</td>
<td>1111</td>
<td></td>
<td></td>
<td>11</td>
</tr>
</tbody>
</table>

134
| Extent to which communication is open | 1111 | 111 | 1 | 11 |
| Extent to which courses are made available | 1111 | 111 | 1 | 11 |
| Extent to which donations are solicited and collected | 1 | 111 | 111 | 1 |
| Extent to which evaluation is comprehensive, not single element | 1111 | 111 | 1 | 11 |
| Extent to which evaluation is ongoing | 1111 | 111 | 1 | 11 |
| Extent to which general public accepts/supports program | 1111 | 11111 | 11 |
| Extent to which general public is educated about motorcycles | 1111 | 1 | 1 | 11 |
| Extent to which graduates think they are given a better head start | 1111 | 1111 | 11 |
| Extent to which Instructors exceed minimum expectations | 111111 | 1 | 11 |
| Extent to which non-core courses are offered | 1111 | 11 | 1 | 1 |
| Extent to which policymakers’ expectations are exceeded | 1111 | 111 | 1 |
| Extent to which pre-testing/post-testing occurs | 111 | 11 | 1 | 11 |
| Extent to which program participation is fun | 111111 | 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 quality and quantity are raised simultaneously | 1111 | 111 | 11 |
| Extent to which self-evaluation is accomplished and enhanced throughout program | 111 | 11111 | 11 |
| Extent to which smaller markets are served | 1111 | 111 | 1 | 11 |
| Extent to which sponsors are provided resources | 1111 | 11 | 1 | 1 |
| Extent to which stakeholders can ask “Why” and “So What” | 1111 | 1111 | 1 |
| Feedback from professional educators is solicited and documented | 11 | 11 | 1 | 11 |
| Focusing on process over focusing on results | 1111 | 1 |
| Following national guidelines | 11111 | 11 |
| Formal annual reports are completed | 111111 | 11 |
| Formal evaluation of program administrator | 1111 | 11 | 1 |

135
| Category                                                                 | Frequency | N/A | Frequency | N/A | Frequency | N/A | Frequency | N/A | Frequency | N/A | Frequency | N/A | Frequency | N/A | Frequency | N/A | Frequency | N/A | Frequency | N/A | Frequency | N/A | Frequency | N/A | Frequency | N/A | Frequency | N/A |
|--------------------------------------------------------------------------|-----------|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|-----------|-----|
| Frequency of Instructor observations                                      | 111       |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Frequency of Instructor updates                                          | 11        |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Frequency of meeting with Instructors                                    | 11        |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Frequency of personnel turnover                                          | 11        |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Having adequate budget to meet expectations                               | 11111     | 1   |           | 1   |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Having adequate number of motorcycles                                     | 11111     | 1   |           | 1   |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Having adequate storage                                                   | 111       | 111  |           | 11  |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Having democratic, open communication                                     | 111       | 111  |           | 11  |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Having incentives for course enrollment                                   | 111       | 111  |           | 11  |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Having skill test waiver for course completion                            | 1111      | 11  |           | 1   |           | 1111 |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Having standardized ranges                                                | 11111     | 11  |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Having website to provide student/graduate feedback                       | 11        | 111  |           | 1   |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Increase in licensed riders                                               | 1111      | 111  |           | 1   |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Increase in motorcycle licensure                                          | 11111     | 11   |           | 1   |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Increase in ratio of license holders/permit holders                       | 11111     | 11   |           | 1   |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Increased learning by participants                                       | 111111    |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Increased skill of participants                                          | 11111     | 11   |           | 1   |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Input of motorcycle groups                                                | 111       | 111  |           | 1   |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Instructor longevity in program                                          | 11111     | 11111 |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Instructors CPR/1st aid certified                                        | 11        | 11111 |           | 1   |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Involvement of professional educators in program                          | 111       | 111  |           | 1   |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |
| Lack of complaints                                                        | 11        | 111  |           | 11  |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |           |     |

136
<table>
<thead>
<tr>
<th></th>
<th>11111</th>
<th>11</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of communication within program</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level of volunteering by Instructors</td>
<td>1</td>
<td>11111</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Low number of no-shows</td>
<td>11</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintenance condition of motorcycles</td>
<td>11111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Making it easy to obtain license for qualified riders</td>
<td>11111</td>
<td>11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Membership in professional organizations</td>
<td>11111</td>
<td>11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Minimal Instructor burnout</td>
<td>11111</td>
<td>111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mirroring of MSF guidelines/recommendations</td>
<td>11111</td>
<td>111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network of communication among Instructors</td>
<td>11</td>
<td>11111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of administrative meetings</td>
<td>11</td>
<td>111</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Number of conferences attended by personnel</td>
<td>11111</td>
<td>11</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Number of full-time administrative personnel</td>
<td>11111</td>
<td>11</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Number of full-time people in program</td>
<td>11111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of motorcycle fatalities</td>
<td>11</td>
<td>111</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Number of potential enrollees turned away</td>
<td>11</td>
<td>111</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Number of site visits per year</td>
<td>11111</td>
<td>11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Number/frequency of accidents in courses</td>
<td>11111</td>
<td>111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numbers trained</td>
<td>11</td>
<td>11111</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>On-line registration availability</td>
<td>11111</td>
<td>11</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Participant cost for program</td>
<td>11111</td>
<td>11</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Pass/fail rate in courses</td>
<td>11111</td>
<td>11111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>People skills of administrators and Instructors</td>
<td>11111</td>
<td>11</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>People’s first impression</td>
<td>11111</td>
<td>11</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Percent of course participants that get licenses</td>
<td>11111</td>
<td>11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Percent of riders reached</td>
<td>11111</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Philosophy and interest of program administrator</td>
<td>11111</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Priority of agency in which program is conducted</td>
<td>11111</td>
<td>11</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Professional development beyond classroom/range instruction</td>
<td>111</td>
<td>111</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Program structure is flat and non-hierarchical</td>
<td>1</td>
<td>111</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Providing re-tests without having entire course repeated</td>
<td>1</td>
<td>1</td>
<td>1111</td>
<td>1</td>
</tr>
<tr>
<td>Provisions for professional growth</td>
<td>111</td>
<td>111</td>
<td>1</td>
<td>111</td>
</tr>
<tr>
<td>Quality of fit in state traffic safety system</td>
<td>1</td>
<td>1111</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Quality of instruction and Instructors</td>
<td>11111</td>
<td>11</td>
<td></td>
<td>11111</td>
</tr>
<tr>
<td>Quality of Instructor updates</td>
<td>111111</td>
<td>1</td>
<td></td>
<td>1111</td>
</tr>
<tr>
<td>Quality of interaction between Instructors and students</td>
<td>11111</td>
<td>11</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Quality of original needs assessment</td>
<td>11111</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Quality of promotional activities</td>
<td>11</td>
<td>111</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Quality of record keeping</td>
<td>111</td>
<td>1111</td>
<td></td>
<td>111</td>
</tr>
<tr>
<td>Quality of screening Instructor candidates</td>
<td>111</td>
<td>1</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Ratio of course participants to licensed riders</td>
<td>1</td>
<td>11111</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Recognition of people involved in program</td>
<td>1</td>
<td>11111</td>
<td>1</td>
<td>111</td>
</tr>
<tr>
<td>Reduced emphasis on crashes and violations as an indicator</td>
<td>1</td>
<td>111</td>
<td>111</td>
<td>11</td>
</tr>
<tr>
<td>Reduces amount of illegal riding</td>
<td>11</td>
<td>111</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Requiring that student evaluations be completed</td>
<td>11</td>
<td>11111</td>
<td>1</td>
<td>11111</td>
</tr>
<tr>
<td>Sense of community within program stakeholders</td>
<td>11</td>
<td>1111</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Separating &quot;assessment of program&quot; from &quot;assessment of course participants&quot;</td>
<td>1</td>
<td>1111</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Size of policies and procedures manual (larger being better)</td>
<td>11</td>
<td>1111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source of funding</td>
<td>11</td>
<td>1111</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Stakeholder involvement in decision-making</td>
<td>11</td>
<td>1111</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>State program compliance with the letter of the law</td>
<td>1111</td>
<td>1</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Strength of local Instructors and coordinators</td>
<td>111</td>
<td>1111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sufficient number of motorcycles</td>
<td>111</td>
<td>1111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feature</td>
<td>1</td>
<td>11</td>
<td>111</td>
<td>1111</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>---</td>
<td>----</td>
<td>-----</td>
<td>------</td>
</tr>
<tr>
<td>Support from superiors</td>
<td>1</td>
<td>11</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Tiered instructor system</td>
<td>1</td>
<td>11</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Time lines for Instructor follow up</td>
<td>11</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Training itself is safe</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Training site coverage in state</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Tuition is low</td>
<td>111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Uniformity of course reporting</td>
<td>1111</td>
<td>111</td>
<td>1111</td>
<td>1</td>
</tr>
<tr>
<td>Use of 3rd party testing</td>
<td>111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Use of advisory board</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Use of credit cards for tuition payment</td>
<td>111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Use of formal research</td>
<td>111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Use of internet</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Use of outside evaluators</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Use of peer critiques</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Use of standardized assessments</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Use of student and graduate evaluations</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Use of website registration</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Using background checks on Instructors</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Videotaping is used for Instructor development</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Visibility of program within the state</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Vision of administrators is growth-oriented versus controlling-and-monitoring-oriented</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Wait time for classes</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Wait time for classes less than 90 days</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Way in which re-scheduling is accomplished</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td>Way in which re-tests are conducted</td>
<td>1111</td>
<td>111</td>
<td>111</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1111</td>
<td>111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>------</td>
<td>-----</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Way problems are addressed</td>
<td>1111</td>
<td>111</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Way staff is managed</td>
<td>11</td>
<td>1111</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Way students are dismissed</td>
<td>1111</td>
<td>1111</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Where program situated in state system</td>
<td>1111</td>
<td>1111</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Written and verbal testimonials are captured/document</td>
<td>1</td>
<td>1111</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Appendix B

Evaluation Criteria: Round Two with Frequency of Responses
### Evaluation Criteria: Round Two with Frequency of Responses

<table>
<thead>
<tr>
<th>Criterion</th>
<th>More Value</th>
<th>Of Value</th>
<th>Less Value</th>
<th>Criterion not understood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to remedy/remove Problems within program</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Accomplishment of pre-determined goals</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Amount of funding</td>
<td>5</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Capturing and acting on input from Instructors</td>
<td>5</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Clear communication as the &quot;measuring stick&quot; for program</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Course student evaluation results</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Dedication of the people involved</td>
<td>5</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Degree of emphasis on service function</td>
<td>5</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Degree of professional development among Instructors</td>
<td>5</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Degree of public support</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Degree of stakeholder feedback</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Degree of trust within program</td>
<td>5</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Degree to which administering agency understands program</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Dollars spent on quality assurance</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Extent of constant learning within program</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Extent to which communication is open</td>
<td>5</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Extent to which general public is educated about motorcycles</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Extent to which Instructors exceed minimum expectations</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Extent to which program participation is fun</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Extent to which quality assurance measures are implemented</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Extent to which quantity and quality are raised simultaneously</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Following national guidelines</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Frequency of quality assurance visits</td>
<td>1</td>
<td>1111</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Frequency of technical assistance visits</td>
<td>111</td>
<td>111</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Having adequate budget to meet expectations</td>
<td>11111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Having adequate number of motorcycles</td>
<td>11111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Having skill test waiver for course completion</td>
<td>11</td>
<td>111</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Increase in licensed riders</td>
<td>11</td>
<td>1111</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Increase in motorcycle licensure</td>
<td>11</td>
<td>111</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>Increase in ratio of license holders/permit holders</td>
<td>1</td>
<td>111</td>
<td>111</td>
<td></td>
</tr>
<tr>
<td>Increased learning by participants</td>
<td>11111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Increased skill of participants</td>
<td>111</td>
<td>111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level of communication within program</td>
<td>11111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maintenance condition of motorcycles</td>
<td>11111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>People skills of administrators and instructors</td>
<td>11111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>People's first impression</td>
<td>111</td>
<td>1111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Priority of agency in which program is conducted</td>
<td>1111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of instruction and Instructors</td>
<td>1111111</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of Instructor updates</td>
<td>1111111</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of interaction between Instructors and students</td>
<td>11111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>State program compliance with the letter of the law</td>
<td>111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Support from superiors</td>
<td>1111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training itself is safe</td>
<td>11111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training site coverage in state</td>
<td>1111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of peer critiques</td>
<td>111</td>
<td>111</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Way problems are addressed</td>
<td>1111</td>
<td>11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

143
Appendix C

Administrator Letter of Invitation
March 21, 2000

Dear:

I am writing to ask for your assistance in a research study titled “Criteria for Program Evaluation in Motorcycle Safety Rider Education and Training.” Your primary involvement would be to participate in a one-to-two hour interview to provide your perspectives as a state administrator. The interview will be audio taped for transcription purposes, but information in the study records will be kept strictly confidential. There should be no risks for you in this study.

I am a faculty member in the Department of Loss Prevention and Safety at Eastern Kentucky University. This study is being performed in connection with my dissertation in pursuit of a doctorate at the University of Tennessee. The purpose of this research is to develop criteria for program evaluation. You are one of eight state administrators in the southeast region of the United States being asked to participate, and you were identified using the State Motorcycle Safety Administrators “State Program Motorcycle Survey.”

Enclosed you will find two copies of an informed consent statement, one for your records and one to be signed and returned to me in the enclosed envelope. Also enclosed is a listing of the core interview questions for your review.

When I receive the signed informed consent statement, I will contact you to schedule an interview at a time and place convenient for you. At the time of the interview, I would ask that you provide, if possible, a copy of original documentation related to program implementation. Your participation is voluntary and you may stop at any time. Questions about the study should be directed to me at (606) 622-2236 or e-mail: lpsochs@acs.eku.edu.

Thank you for your consideration in participating in this study.

Regards,

Ray Ochs
Coordinator

Enclosures
August 25, 2000

Dear,

Thank you for recently completing the rating of criteria for program evaluation in motorcycle safety programs.

Enclosed you will find a list of criteria that were rated as having “more value” than others by at least half of the administrators that completed the rating. Please complete the form and return it to me in the enclosed pre-addressed, stamped envelope. Should you have any questions, feel free to contact me at 859-622-1012 or 800-396-3234.

There are directions are on the form, but basically it requires a checkmark for each of the criteria in one of three categories (more value, of value, less value), and a column to check if a criterion is not understood. It might be wise to look at several criteria before marking the form to get an idea of the range of criteria as ideally there would be an equal number of criteria in each of the three rating categories.

Thank you for your time and efforts in this project.

Regards,

Ray Ochs
Appendix E

Form B: Human Subjects Form
THE UNIVERSITY OF TENNESSEE, KNOXVILLE

Application for Review of Research Involving Human Subjects

I. IDENTIFICATION OF PROJECT

1. Principal Investigator (PI) or Co-Principal Investigators (Co-PI)
   Raymond J. Ochs
   329 Boone Trail
   Richmond, Kentucky 40475
   Home Telephone: (606) 625-0342
   Office Telephone: (606) 622-1012
   email address: lpsochs@acs.eku.edu

2. Project Classification: Dissertation

3. Project Title: “Criteria for Program Evaluation in Motorcycle Safety Rider Education and Training”

4. Starting Date: Upon IRB Approval

5. Estimated Completion Date: Fall, 2000

6. External Funding: N/A

II. PROJECT OBJECTIVES:

The objective of the project is to develop criteria for program evaluation in motorcycle safety rider education and training programs from the perspective of program administrators. Program administrators from eight states will be interviewed. The research is being conducted in partial fulfillment of a doctoral dissertation.
III. DESCRIPTION AND SOURCE OF RESEARCH PARTICIPANTS

The participants will be state administrators of motorcycle safety rider education and training programs. Selections will be based on identifying the person responsible for the day-to-day program activities in the eight states that form the southeast region of the National Highway and Traffic Safety Administration. The states are Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, and Tennessee. The subjects will be interviewed for approximately one to two hours (see attachment #1 for primary interview questions). Participants' involvement in the project will span approximately six months from the date of project inception.

IV. METHODS AND PROCEDURES

A program administrator from each state will be invited by letter to participate in the study. The researcher will ask for an interview which will be held in the participant's state at a location such as an office or reasonable off-site location of the administrator's choice. The letter will identify the purpose of the study, the research procedures, and how the administrator was identified.

The researcher will emphasize that the interview is completely voluntary, and any questions she or he may have will be answered, that identities and state-specific information will be kept anonymous through the use of pseudonyms, and that the participant may withdraw from the study at any time without penalty. A written informed consent form will be utilized (see attachment #2). The interview will be open-ended, and will be audio-taped and transcribed. The reason to tape and transcribe the interviews is to be able to capture the details of the interviews for the purpose of analyzing the data. The tapes will be destroyed after transcription by smashing the plastic housing and depositing the broken plastic pieces and tape in a trash receptacle. The transcriptions will be maintained under lock and key in the researcher's office for a period of one year after completion of the project. Only the researcher and transcriptionist will have access to the data. The transcriptionist will complete a statement of confidentiality.

The data will be analyzed and interpreted using a constant comparative method. This involves constructing categories and subcategories from units of data that have something in common. Data will be analyzed as it is being collected, and again as the transcripts are reviewed.

V. SPECIFIC RISKS AND PROTECTION MEASURES

The risks of harm anticipated in the proposed research are not greater, considering probability and magnitude, than those ordinarily encountered in daily life or during the performance of routine duties. Pseudonyms are being used in place of respondent's names, and state-specific information will be purged from the transcripts. Only the researcher and transcriptionist will have access to the data, and the files will be secure in the researcher's office under lock and key. The transcriptionist will complete a statement of confidentiality.

VI. BENEFITS
The risks are minimal, and the benefit is to contribute to the knowledge in the field of motorcycle safety rider education and training programs.

VII. METHODS FOR OBTAINING "INFORMED CONSENT" FROM PARTICIPANTS

To obtain legally effective informed consent from participants, a written informed consent form containing all the basic elements of informed consent will be utilized. This form will be signed by the participant and an extra copy will be provided for each participant. The informed consent form will be mailed in a manner to allow prospective participants a minimum of two weeks to consider participation. The language used in the informed consent form will be understandable to the prospective participants. Each prospective participant will be able to contact the researcher to ask any questions or request further clarification of the procedures or involvement in the study. Signed consent forms will be stored in the researcher's office at the Traffic Safety Institute at Eastern Kentucky University, and will be available to University of Tennessee, Knoxville personnel. Signed informed consent documents will be kept for the duration of the project and for at least three years thereafter in the researcher's office.

VIII. QUALIFICATIONS OF THE INVESTIGATOR

The investigator is a doctoral candidate in Psychoeducational Studies at The University of Tennessee, Knoxville. He is employed by Eastern Kentucky University as Coordinator and Associate Professor at the Traffic Safety Institute. He is a chief instructor for the Motorcycle Safety Foundation and a member of the National Association of State Motorcycle Safety Administrators.

IX. FACILITIES AND EQUIPMENT TO BE USED IN THE RESEARCH

The facilities and equipment to be used in this research includes the investigator's office and office equipment, including computer and word processing software. Locations of the interviews will be at a location in the participant's state. Metal file cabinets in the researcher's office will be used for storage of the data.

X. RESPONSIBILITY OF THE PRINCIPAL/CO-PRINCIPAL INVESTIGATOR(S)

By compliance with the policies established by the Institutional Review Board of The University of Tennessee, Knoxville, the principal investigator(s) subscribe to the principles stated in "The Belmont Report" and standards of professional ethics in all research, development, and related activities involving human participants under the auspices of The University of Tennessee, Knoxville. The principal investigator(s) further agree that:

1. Approval will be obtained from the Institutional Review Board prior to instituting any change in this research project.
2. Development of any unexpected risks will be immediately reported to the Compliances Section.
3. An annual review and progress report (Form R) will be completed and submitted when requested by the Institutional Review Board.
4. Signed informed consent documents will be kept for the duration of the project and for at least three years thereafter at a location approved by the Institutional Review Board.
XI. SIGNATURES

Principal Investigator

Signature ___________________________ Date ________________

Student Advisor (if any)

Signature ___________________________ Date __________________

XI. DEPARTMENT REVIEW AND APPROVAL

The IRB department review committee has reviewed and approved the application described above. The DRC recommends that this application be reviewed as:

[ ] Expedited Review – Category(ies): ______________

OR

[ ] Full IRB Review

Chair, DRC ____________________________

Signature ___________________________ Date ________________

Department Head _______________________

Signature ___________________________ Date ________________

Protocol sent to Compliance Section for final approval on (Date) ______________________

Approved: Compliance Section

Office of Research

404 Andy Holt Tower

Signature ___________________________ Date ________________
Appendix F

Informed Consent Statement
INFORMED CONSENT FORM

"Criteria for Program Evaluation in Motorcycle Safety Rider Education and Training"

You are invited to participate in a research study. The purpose of the study is to develop criteria for program evaluation in motorcycle safety rider education and training programs from the perspective of program administrators.

INFORMATION

Certain procedures will be utilized in conducting and completing this study. A program administrator for each of the eight states in the southeast region of the National Highway Traffic Safety Administration will be invited to participate in the study. An interview, lasting from one to two hours, will be conducted at a location determined by the program administrator. Original documentation related to program implementation will be requested. The interview will be audio-taped and transcribed. The tapes will be destroyed after transcription. The same primary questions will be asked of each administrator, and the questions will be mailed to the administrators prior to the interview. Upon completion of the interviews, the researcher will develop criteria for program evaluation. These criteria will be sent to each interviewee for comments regarding accuracy, clarity, and utility. The amended criteria developed from this input will again be mailed to each interviewee for additional comments. From this information, a final set of criteria for program evaluation will be determined.

The amount of time required of participants is approximately two hours for the interview process, and one hour of additional time to review the criteria developed for program evaluation. Participants' involvement in the project will span approximately six months from the date of project inception.

Participation in the study is completely voluntary, and any questions participants may have will be answered. Identities and state-specific information will be kept anonymous through the use of pseudonyms. Withdrawal from the study can occur at any time.

RISKS

There are no foreseeable risks involved in this research beyond what occur during the performance of routine duties.

[Participant's initials]
BENEFITS

This research may lead to a new and viable approach to program evaluation in state motorcycle safety programs. Additionally, new information may be discovered regarding program evaluation from state administrators' perspective. Findings are likely to add to the knowledge base of motorcycle safety rider education and training programs.

CONFIDENTIALITY

Information in the study records will be kept confidential. Data will be stored securely and will be made available only to persons conducting the study. No reference will be made in oral or written reports which could link participants to the study.

CONTACT

If you have any questions at any time about the study or the procedures, (or you experience adverse effects as a result of participating in this study,) you may contact the researcher, Ray Ochs, at 250 Stratton, Eastern Kentucky University, Richmond, Kentucky 40475, and (606) 622-1012. If you have questions about your rights as a participant, contact the Compliance Section of the Office of Research at (423) 974-3466.

PARTICIPATION

Your participation in this study is voluntary; you may decline to participate without penalty. If you decide to participate, you may withdraw from the study at anytime without penalty and without loss of benefits to which you are otherwise entitled. If you withdraw from the study before data collection is completed your data will be returned to you or destroyed.

CONSENT

I have read the above information and agree to participate in this study. I have received a copy of this form.

Participant's signature (print)

Participant's signature

Date
VITA

Raymond J. Ochs is associate professor and coordinator at the Traffic Safety Institute at Eastern Kentucky University in Richmond, Kentucky. He has been in that position since 1976. He was previously employed as Instructor at the Traffic Safety Center at East Carolina University, and also held a faculty position as Lecturer at Appalachian State University. Prior to serving as a graduate assistant at Indiana State University’s Driver Education Instructional Demonstration Center in 1971-72, Mr. Ochs taught high school health, physical education and driver education at LeBlond High School in St. Joseph, Missouri.

Mr. Ochs received his undergraduate degree in Secondary Education from Northwest Missouri State University in 1970, and completed the Master’s degree in Health and Safety at Indiana State University in 1972. He has completed additional course work at the University of Maryland and Eastern Kentucky University.

Mr. Ochs frequently consults with agencies and organizations with an interest in traffic safety and motorcycle safety. He is licensed by the Specialty Vehicle Institute of America as a Chief Instructor Trainer and is a certified Chief Instructor for the Motorcycle Safety Foundation. He has been certified as an instructor-trainer for the General Motors Advanced Driving Program and the National Safety Council’s Defensive Driving Program.